# Big Data Analytics in Intelligent Transportation Systems A Survey
智能交通系统中的大数据分析调查



## Transportation Systems: A Survey
## 运输系统：一项调查

Li Zhu, Fei Richard Yu, *Fellow, IEEE*, Yige Wang, Bin Ning, *Fellow, IEEE*, and Tao Tang



Abstract— Big data is becoming a research focus in intelligent transportation systems (ITS), which can be seen in many projects around the world. Intelligent transportation systems will produce a large amount of data. The produced big data will have profound impacts on the design and application of intelligent transportation systems, which makes ITS safer, more efficient, and profitable. Studying big data analytics in ITS is a flourishing field. This paper first reviews the history and characteristics of big data and intelligent transportation systems. The framework of conducting big data analytics in ITS is discussed next, where the data source and collection methods, data analytics methods and platforms, and big data analytics application categories are summarized. Several case studies of big data analytics applications in intelligent transportation systems, including road traffic accidents analysis, road traffic flow prediction, public transportation service plan, personal travel route plan, rail transportation management and control, and assets maintenance are introduced. Finally, this paper discusses some open challenges of using big data analytics in ITS.
摘要 - 大数据正成为智能交通系统（ITS）的研究热点，可以在世界各地的许多项目中看到。智能交通系统将产生大量数据。生成的大数据将对智能交通系统的设计和应用产生深远的影响，使ITS更安全，更高效，更有效。在ITS中研究大数据分析是一个很好的领域。本文首先回顾了大数据和智能交通系统的历史和特点。接下来讨论在ITS中进行大数据分析的框架，其中总结了数据源和收集方法，数据分析方法和平台以及大数据分析应用类别。介绍了智能交通系统中大数据分析应用的几个案例研究，包括道路交通事故分析，道路交通流量预测，公共交通服务计划，个人旅行路线规划，铁路运输管理和控制以及资产维护。最后，本文讨论了在ITS中使用大数据分析的一些开放性挑战。

Index Terms— Big data analytics, intelligent transportation systems (ITS), machine learning, transportation.

索引条款 - 大数据分析，智能交通系统（ITS），机器学习，运输。

# I. INTRODUCTION
# 一，导言

R ECENTLY, Big Data has become a hot topic in bothacademia and industry. It represents large and complex data sets obtained from all kinds of sources. Many of the most popular data process techniques contain Big Data techniques, including data mining, machine learning, artificial intelligence, data fusion, social networks and so on [1]. Many people use Big Data analytics in various fields, and have achieved great success [2]. For example, in business field, some enterprises use Big Data to understand the consumer behavior more accurately so as to optimize the product price, improve

R ECENTLY，大数据已成为bothacademia和行业的热门话题。它代表从各种来源获得的大而复杂的数据集。许多最流行的数据处理技术都包含大数据技术，包括数据挖掘，机器学习，人工智能，数据融合，社交网络等[1]。许多人在各个领域使用大数据分析，并取得了巨大的成功[2]。例如，在商业领域，一些企业使用大数据更准确地了解消费者行为，从而优化产品价格，改善

Manuscript received August 1, 2017; revised January 14, 2018; accepted February 24, 2018. Date of publication April 23, 2018; date of current version December 21, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61603026, in part by the Beijing Natural Science Foundation under Grant L171004, in part by the Technological Research and Development Program of China Railway Corporation under Grant 2016X008-B, in part by the State Key Laboratory of Rail Traffic Control and Safety under Grant RCS2017ZT006, and Project KIE017001531, and in part by the Beijing Key Laboratory of Urban Rail Transit Automation and Control. The authors declare that there is no conflict of interest regarding the publication of this paper. The Associate Editor for this paper was J. E. Naranjo. (Corresponding author: Fei Richard Yu.) L. Zhu, Y. Wang, B. Ning, and T. Tang are with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China (e-mail: lizhu@bjtu.edu.cn; 15120287@bjtu.edu.cn; bning@bjtu.edu.cn; ttang@bjtu.edu.cn).
手稿于2017年8月1日收到;修订于2018年1月14日; 2018年2月24日接受。出版日期2018年4月23日;当前版本的日期2018年12月21日。这项工作部分得到了国家自然科学基金项目61603026的资助，部分得到了北京市自然科学基金项目L171004的资助，部分得到了中国铁路总公司技术研究发展计划项目2016X008-B的资助。部分由RCS2017ZT006授权的轨道交通控制与安全国家重点实验室和KIE017001531项目组成，部分由北京市轨道交通自动化与控制重点实验室完成。作者声明，对于本文的发表没有任何利益冲突。本文的副主编是J. E. Naranjo。（通讯作者：Fei Richard Yu。）L。Zhu，Y。Wang，B。Ning和T. Tang在北京交通大学轨道交通控制与安全国家重点实验室，北京100044（电子邮件：lizhu@bjtu.edu.cn; 15120287@bjtu.edu.cn; bning@bjtu.edu.cn; ttang@bjtu.edu.cn）。

F. R. Yu is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada (e-mail: richard.yu@ carleton.ca).
F. R. Yu在加拿大ON K1S 5B6，渥太华卡尔顿大学系统与计算机工程系（电子邮件：richard.yu @ carleton.ca）。

Digital Object Identifier 10.1109/TITS.2018.2815678
数字对象标识符10.1109 / TITS.2018.2815678

383

operational efficiency and reduce personnel costs [3]. In social network field [3], through Big Data analytics of instant messaging, online social networking, microblog and sharing space, some companies such as Facebook, Twitter and Linkedin can understand the user's current behavior, social connections and rules of social behavior, and then promote some products. In health care field, by processing, and querying of health care data, doctors can analyze the pathogenic characteristics, assessment of the patient's physique so as to develop more humane treatment plans and suggestions and reduce incidence of patients [4]. In smart grid field, via the analysis of smart grid data, grid operators can know which parts of the electricity load and power frequency are too high, and even can diagnose which lines are in failure state. The results of these data analysis can be contributed to the upgrading of the electrical grid, renovation and maintenance work [5]. With successful application of Big Data analytics in so many fields, intelligent transportation systems also start looking at Big Data with great interests.
运营效率和降低人员成本[3]。在社交网络领域[3]，通过即时消息，在线社交网络，微博和共享空间的大数据分析，Facebook，Twitter和Linkedin等一些公司可以了解用户当前的行为，社交关系和社交行为规则，以及然后推广一些产品。在医疗保健领域，通过处理和查询医疗保健数据，医生可以分析病原体特征，评估患者的体质，从而制定更加人性化的治疗方案和建议，降低患者的发病率[4]。在智能电网领域，通过对智能电网数据的分析，电网运营商可以知道电力负荷和电力频率的哪些部分过高，甚至可以诊断哪些线路处于故障状态。这些数据分析的结果可以促进电网的升级，改造和维护工作[5]。随着大数据分析在如此多的领域中的成功应用，智能交通系统也开始以极大的兴趣关注大数据。

Intelligent transportation systems (ITS) have been developed since the beginning of 1970s. It is the future direction of the transportation system. ITS incorporate advanced technologies which include electronic sensor technologies, data transmission technologies, and intelligent control technologies into the transportation systems [6]. The purpose of ITS is to provide better services for drivers and riders in transportation systems [7]–[9].
自20世纪70年代初以来，智能交通系统（ITS）得到了发展。这是交通系统的未来发展方向。ITS将先进的技术，包括电子传感器技术，数据传输技术和智能控制技术融入运输系统[6]。ITS的目的是为运输系统中的驾驶员和乘客提供更好的服务[7] - [9]。

In ITS, data can be obtained from diverse sources, such as smart card, GPS, sensors, video detector, social medias, and so on. Using accurate and effective data analytics of seemingly disorganized data can provide better service for ITS [10], [11]. With the development of ITS, the amount of data generated in ITS is developing from Trillionbyte level to Petabyte. Given such amount of data, traditional data processing systems are inefficient, and cannot meet the data analytics requirement. This is because they do not foresee the rapid growth of data amount and complexity.
在ITS中，数据可以从各种来源获得，例如智能卡，GPS，传感器，视频检测器，社交媒体等。对看似无组织的数据进行准确有效的数据分析可以为ITS提供更好的服务[10]，[11]。随着ITS的发展，ITS中生成的数据量正在从万亿字节级发展到Petabyte级。鉴于数据量如此之大，传统的数据处理系统效率低下，无法满足数据分析的要求。这是因为他们没有预见到数据量和复杂性的快速增长。

Big Data analytics provides ITS a new technical method. ITS can benefit from Big Data analytics in the following aspects.
大数据分析为ITS提供了一种新的技术方法。ITS可以从以下几个方面受益于大数据分析。

1. Vast amounts of diverse and complex data generated in ITS can be handled by Big Data analytics. Big Data analytics has resolved three problems: data storage, data analysis and data management. Big Data platforms such as Apache Hadoop and Spark are capable to processing massive amounts of data, and they have been widely used in academia and industry [12], [13].
1.大数据分析可以处理ITS中生成的大量不同且复杂的数据。大数据分析解决了三个问题：数据存储，数据分析和数据管理。Apache Hadoop和Spark等大数据平台能够处理大量数据，并且已广泛应用于学术界和工业界[12]，[13]。

2. Big Data analytics can improve the ITS operation efficiency. Many subsystems in ITSs need to handle large amount of data to give information or provide decision to manage traffic. Through fast data collection and analysis of current and historical massive traffic data, traffic management department can predict traffic flow in real time. Public transportation Big Data analytics can help management department to learn the riders journey patterns in the transportation network, which can be used for better public transportation service planning. Big Data analytics of transportation APP developers can help the users to reach their destination in a most suitable route and with the shortest possible time.
2.大数据分析可以提高ITS运营效率。ITS中的许多子系统需要处理大量数据以提供信息或提供管理流量的决策。通过快速数据收集和当前和历史海量交通数据的分析，交通管理部门可以实时预测交通流量。公共交通大数据分析可以帮助管理部门了解交通网络中的乘客旅程模式，这可以用于更好的公共交通服务规划。运输APP开发人员的大数据分析可以帮助用户以最合适的路线以最短的时间到达目的地。

3. Big Data analytics can improve the ITS safety level. Using advanced sensor and detection techniques, massive amount of real time transportation information can be obtained. Through Big Data analytics, we can effectively predict the occurrence of traffic accident. When accidents happens, or emergency rescue is needed, the real time response capability in the Big Data analytics based system can greatly improve the emergency rescue ability. Big Data analytics can also offer new opportunities to identify assets problems, such as pavement degradation, ballast aging, etc. It can help make maintenance decision in an appropriate time, and prevent the vehicle or infrastructure from being in a failure state.
3.大数据分析可以提高ITS安全水平。使用先进的传感器和检测技术，可以获得大量的实时运输信息。通过大数据分析，我们可以有效地预测交通事故的发生。当事故发生或需要紧急救援时，基于大数据分析的系统中的实时响应能力可以极大地提高应急救援能力。大数据分析还可以提供识别资产问题的新机会，例如路面退化，压载老化等。它可以帮助在适当的时间做出维护决策，并防止车辆或基础设施处于故障状态。

Although applications of Big Data analytics in ITS have the great vision, many critical research issues and significant challenges remain need to be addressed. To the best of our knowledge, a systematic summary of Big Data analytics from data sources and collection methods, data analytics methods and platforms, to Big Data analytics applications in ITS has not been done before. In this survey, we first discuss the sources of Big Data in ITS and how we can collect the generated Big Data. The framework of conducting Big Data analytics in ITS is discussed. We also summarize the data analytics methods and platforms in ITS. Some case studies of Big Data analytics applications in ITS are introduced as well.
虽然大数据分析在ITS中的应用具有很好的视野，但仍有许多关键的研究问题和重大挑战需要解决。据我们所知，从数据源和收集方法，数据分析方法和平台到ITS中的大数据分析应用程序的大数据分析的系统总结以前还没有完成。在本次调查中，我们首先讨论了ITS中大数据的来源以及我们如何收集生成的大数据。讨论了在ITS中进行大数据分析的框架。我们还总结了ITS中的数据分析方法和平台。还介绍了ITS中大数据分析应用的一些案例

研究。

The rest of paper is organized as follows. The architecture of conducting Big Data analytics in ITS is discussed in Section II. Section III summarizes the data source and collection methods. Big Data analytics methods are discussed in Section IV. Section V introduces the cases studies of ITS Big Data analytics applications in details. We present the Big Data analytics platforms in Section VI. Some open challenges of using Big Data analytics in ITS are discussed in Section VII. Finally, We conclude the paper in Section VIII.
其余论文的结构如下。第二部分讨论了在ITS中进行大数据分析的架构。第三节总结了数据来源和收集方法。大数据分析方法在第IV节中讨论。第五节详细介绍了ITS大数据分析应用案例研究。我们在第VI节介绍了大数据分析平台。第VII节讨论了在ITS中使用大数据分析的一些开放性挑战。最后，我们在第八节总结了这篇论文。

## II. THE ARCHITECTURE OF CONDUCTING BIG DATA ANALYTICS IN ITS
## II。在其中进行大数据分析的体系结构

A. Big Data Characteristics in ITS
A. ITS中的大数据特征

Intelligent transportation system incorporates advanced technologies which include electronic sensor technologies, data transmission technologies, and intelligent control technologies into the transportation systems [6]. The purpose of ITS is to provide better services for drivers and riders in transportation systems [7]. According to [7], ITS includes six fundamental components: advanced transportation management systems, advanced traveler information systems, advanced vehicle control systems, business vehicle management, advanced public transportation systems, and advanced urban transportation systems. Literature review [7]–[9] indicates that most of these components are specific to vehicles and road transportation. Therefore, we focus on ITS in-road transportation in this survey paper.
智能交通系统将先进技术融入交通系统[6]，其中包括电子传感器技术，数据传输技术和智能控制技术。ITS的目的是为运输系统中的驾驶员和乘客提供更好的服务[7]。根据[7]，ITS包括六个基本组成部分：先进的运输管理系统，先进的旅行者信息系统，先进的车辆控制系统，商务车辆管理，先进的公共交通系统和先进的城市交通系统。文献综述[7] - [9]表明这些组件中的大多数都是特定于车辆和公路运输的。因此，我们在本调查报告中关注ITS的公路运输。

The data collected by the intelligent transportation systems (ITS) are increasingly complex and are with Big Data features. Big companies including Gartner IBM and Microsoft put forward that that Big Data could be described by three Vs, i.e., volume, variety, and velocity [14], [15].
智能交通系统（ITS）收集的数据越来越复杂，并且具有大数据功能。包括Gartner IBM和微软在内的大公司提出，大数据可以通过三个V来描述，即数量，种类和速度[14]，[15]。

Volume refers to the quantities of data produced by various sources and are still expanding. With the growth of the amount of traffic, and detectors, the volume of data in transportation has increased significantly. In addition, travelers, goods and vehicles generate more data when tracking transponders are used. The data generated from infrastructures, environmental and meteorological monitoring is also increasing as a critical part of transportation data.
数量是指各种来源产生的数据量，并且仍在扩大。随着交通量和探测器数量的增加，运输中的数据量显着增加。

此外，在使用跟踪转发器时，旅行者，货物和车辆会产生更多数据。基础设施，环境和气象监测产生的数据也正在成为运输数据的关键部分。

Variety is mainly focused on all kinds of data produced by detectors, sensors, and even social media. The variety of transport-related data has increased remarkably. For example, modern vehicles can report internal system telemetry in real time and the information of all crew members and passengers.
多样性主要集中在由探测器，传感器甚至社交媒体产生的各种数据上。各种与运输有关的数据显着增加。例如，现代车辆可以实时报告内部系统遥测以及所有机组成员和乘客的信息。

The velocity of data in transportation has increased due to improved communications technologies, increased processing power and speed of monitoring and processing. For example, ticketing and tolling transactions that use smart cards or tags are now immediately reported, whereas paper-based ticketing needs human processing to acquire helpful data from the transactions.
由于通信技术的改进，处理能力的提高以及监测和处理的速度，运输中的数据速度有所提高。例如，现在立即报告使用智能卡或标签的票务和收费交易，而基于纸张的票务需要人工处理以从交易中获取有用的数据。

B. The Architecture of Conducting Big Data Analytics in ITS The architecture of conducting Big Data analytics in ITS is shown in Fig. 1. It can be divided into three layers, which are data collection layer, data analytics layer, and application layer.
B.在ITS中进行大数据分析的体系结构在ITS中进行大数据分析的体系结构如图1所示。它可以分为三层，即数据收集层，数据分析层和应用层。

• Data collection layer: Data collection layer is the basis of the architecture, since it provides the necessary data for the upper layer. The data come from diverse sources such as induction loop detectors, microwave radars, video surveillance, remote sensing, radio frequency identification data, and GPS, etc. Details about collection of Big Data will be introduced in next sections.
•数据收集层：数据收集层是架构的基础，因为它为上层提供了必要的数据。数据来自各种来源，如感应环路探测器，微波雷达，视频监控，遥感，射频识别数据和GPS等。有关大数据收集的详细信息将在下一节中介绍。

• Data analytics layer: Data analytics layer is the core layer of architecture. This layer is primarily to receive data from the data collection layer, and then apply various Big Data analytics approaches and the corresponding platform to complete data storage, management, mining, analysis, and sharing. Details about the Big Data analytics approaches and platform will be introduced in next sections.
•数据分析层：数据分析层是架构的核心层。该层主要用于从数据收集层接收数据，然后应用各种大数据分析方法和相应的平台来完成数据存储，管理，挖掘，分析和共享。有关大数据分析方法和平台的详细信息将在下一节中介绍。

• Application layer: Application layer is the topmost layer in this architecture. It applies the data process results from the data analytics layer in different transportation circumstances, for example, traffic flow prediction, traffic guidance, signal control, and emergency rescue, etc.
•应用层：应用层是此架构中的最顶层。它在不同的运输环境中应用来自数据分析层的数据处理结果，例如，交通流量预测，交通引导，信号控制和紧急救援等。

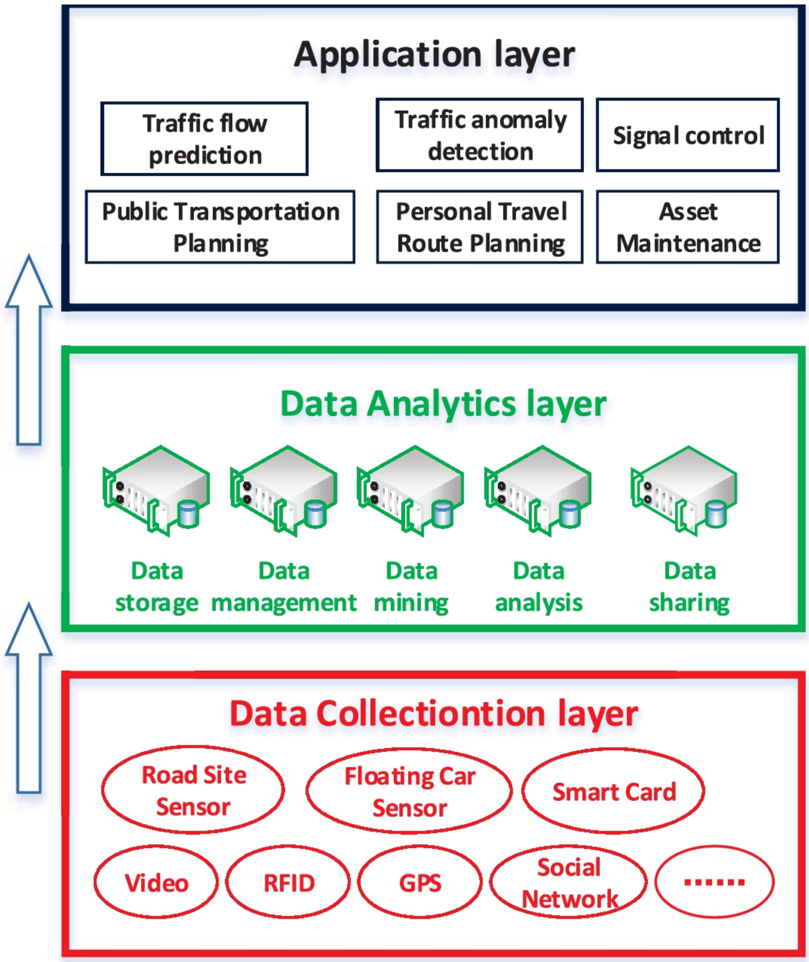



Fig. 1. Architecture of conducting Big Data analytics in ITS.
图1.在ITS中进行大数据分析的架构。

Using advanced data collection techniques, the data collection layer monitors people, vehicles, roads and the environment. The original traffic data which includes structured data, semi-structured and mixed data is transmitted to the data analytics layer via wired or wireless communication. After the data analytics layer receives the original traffic data, it first classifies the data, removes duplicate data, cleans the data and distributes the useful and accurate data in a distributed manner. Then it uses mathematics and engineering theory to extract the hidden information, mainly including descriptive analysis and predictive analysis. Using the analysis results, the application layer can predict the trend of future traffic flow and passengers flow, analyze the traffic accident prone locations, adjust the signal distribution, and implement traffic control to provide decision support for the city management department.
数据采集层使用先进的数据采集技术监控人员，车辆，道路和环境。包括结构化数据，半结构化和混合数据的原始交通数据通过有线或无线通信传输到数据分析层。在数据分析层接收到原始流量数据后，它首先对数据进行分类，删除重复数据，清理数据并以分布式方式分发有用且准确的数据。然后利用数学和工程理论提取隐藏信息，主要包括描述性分析和预测分析。利用分析结果，应用层可以预测未来交通流量和乘客流量的趋势，分析交通事故多发地点，调整信号分布，实施交通控制，为城市管理部门提供决策支持。

## III. BIG DATA COLLECTION IN ITS
## III。其中的大数据收集

People unconsciously participate in the collection, transmission and application of Big Data in ITS. The technology development in ITS has led to an increase in the complexity, diversity and amount of data created and collected from vehicle, and people movements. According to different sources in ITS, Big Data in ITS can be primarily categorized into the following types, and the collected data is illustrated in Table I.

人们无意识地参与了ITS中大数据的收集，传输和应用。ITS中的技术发展导致从车辆和人员流动中创建和收集的数据的复杂性，多样性和数量的增加。根据ITS中的不同来源，ITS中的大数据可以主要分为以下类型，收集的数据如表I所示。

A. Big Data From Smart Cards

A.智能卡的大数据

Automatic Fare Collection (AFC) systems has been widely deployed in urban rail systems, which makes the smart card data become the main data source for investigating the passengers movement patterns [16]–[18]. In AFC systems passengers are required to use smart cards when they take buses or trains. The electronic readers will capture passenger details such as boarding time, OD information, etc., when they touch their smart cards. Smart cards in AFC systems generate huge amount of data records every day in big cities. For instance,

自动票价收集（AFC）系统已广泛应用于城市轨道交通系统，这使得智能卡数据成为调查乘客运动模式的主要数据来源[16] - [18]。在AFC系统中，乘客在乘坐公共汽车或火车时需要使用智能卡。电子阅读器将在触摸智能卡时捕获乘客详细信息，如登机时间，OD信息等。AFC系统中的智能卡每天都会在大城市中生成大量的数据记录。例如，

## TABLE I
## 表I

## BIG DATA IN ITS
## 其中的大数据

| Source | Tools | Data |
|---|---|---|
| Smart Card | Smart Card | OD Flows, Travel Time |
| GPS | GPS | Vehicle Position, Vehicle Density, Vehicle Speed |
| Video | Video Camera | Vehicle Position, Vehicle Speed, Vehicle Density, Vehicle Classification |
| Road Site Sensor | Induction Loops, Road Tubes, Microwave Radar, | Vehicle Position, Vehicle Speed, Vehicle Density, Vehicle |

| | | |
|---|---|---|
| | LIDAR/Infrared Acoustic, Toll Plazas | Classification |
| Floating Car Sensor | License Plate Recognition, Transponders | Travel Time, OD Flows |
| Wide Area Sensor | GPS, Cell phone Tracking, Airborne Sensors | Travel Time, OD Flows |
| Connected and Autonomous Vehicles (CAVs) | Diverse Sensors | Coordinate, speed, acceleration, safety data, |
| Passive Collection | Social Media, Mobile Phone Data | Travel Time, OD Flows |
| Other Sources | Smart Grid, Smart Meters, Cellular Service, Dedicated Tests | Electric and Energy Consumption, Location, Channel Data |

Transportation for London (TfL) collects smart card data from 8 million trips every day at London metro stations.
伦敦交通局（伦敦交通局）每天从伦敦地铁站的800万次旅行中收集智能卡数据。

Substantial work has been done to use smart card data to study the spatial and temporal patterns of public transportation passenger travel behaviour [19]–[22]. Due to its potential capacity of offering comprehensive spatial-temporal information on travel behaviour [17], [21], smart card data is becoming a significant component of public transportation services planning and management.
已经做了大量工作来使用智能卡数据来研究公共交通乘客出行行为的空间和时间模式[19] - [22]。由于其具有提供旅行行为综合时空信息的潜力[17]，[21]，智能卡数据正成为公共交通服务规划和管理的重要组成部分。

B. Big Data From GPS
B.来自GPS的大数据

GPS is the most popular tool for location tracking. Traffic data can be collected more efficiently and safely with location tracking via GPS. Combining geographic information system (GIS) or other map displaying technologies, GPS provides a promising tool for data collection, and the collected data can be used for addressing many traffic issues, such as travel mode detection [23], [24], travel delay measurement [25] and traffic monitoring [26].
GPS是最受欢迎的位置跟踪工具。通过GPS进行位置跟踪，可以更有效，更安全地收集交通数据。结合地理信息系统（GIS）或其他地图显示技术，GPS提供了一种有前途的数据收集工具，收集的数据可用于解决许多交通问题，如旅行模式检测[23]，[24]，旅行延误测量[25]和交通监测[26]。

C. Big Data From Videos

C.视频中的大数据

Video cameras are widely deployed in ITS. As demonstrated in advanced traffic management systems (ATMS), video image detection systems (VIDS) are good alternatives compared with conventional sensors for tasks like vehicle identification and traffic flow detection. One advantage of VIDS is the low cost [27]. Freeway imaging sensors that use massive video data have been successfully deployed to carry out incident detection and have shown high accuracy in certain circumstance [28]. Apart from general traffic management [29], transportation engineers and planners that collect more accurate vehicle video data can improve the image process system so as to be better at making general transportation demand regarding vehicle emission models.
摄像机广泛部署在ITS中。如先进的交通管理系统（ATMS）所示，与传统的传感器相比，视频图像检测系统（VIDS）是车辆识别和交通流检测等任务的理想选择。 VIDS的一个优点是成本低[27]。已成功部署使用海量视频数据的高速公路成像传感器进行事件检测，并在某些情况下显示出高精度[28]。除了一般的交通管理[29]，收集更准确的车辆视频数据的运输工程师和规划人员可以改进图像处理系统，以便更好地制定关于车辆排放模型的一般运输需求。

D. Big Data From Sensors

D.来自传感器的大数据

Sensor equipment installed in ITS is used to collect data such as vehicle speeds, vehicle density, traffic flows, and trip times. Traditional on-road sensors, (e.g., infrared and microwave detectors), have been evolving to obtain, compute and transfer traffic data [30]. As presented in [30], data collection from sensors can be divided into three sources: roadside data, floating car data, and wide area data [31].
ITS中安装的传感器设备用于收集车辆速度，车辆密度，交通流量和行程时间等数据。传统的道路传感器（例如，红外和微波探测器）已经不断发展以获取，计算和传输交通数据[30]。如[30]所示，传感器的数据收集可分为三个来源：路边数据，浮动汽车数据和广域数据[31]。

Roadside data mainly refers to the data collected by sensors located along roadside. Traditional roadside sensors such as inductive magnetic loops, pneumatic road tubes, piezoelectric loops arrays and microwave radars have been used for many years. New generation roadside sensors such as ultrasonic and acoustic sensor systems, magnetometer vehicle detectors, infrared systems, light detection and ranging (LIDAR), and video image processing and detection systems gradually appear with recent advanced technology developments.
路边数据主要是指位于路边的传感器收集的数据。传统的路边传感器，例如感应磁环，气动路管，压电回路阵列和微波雷达已经使用了很多年。新一代路边传感器，如超声波和声学传感器系统，磁力计车辆探测器，红外系统，光探测和测距（LIDAR），以及视频图像处理和探测系统，随着最近的先进技术发展逐渐出现。

Floating car data (FCD) mainly refers to the vehicle mobility data at different locations in ITS, where customized detectors are embedded in vehicles [32]. Some onboard sensors provide confident and efficient information for travel route selection and estimations. With developments of vehicle sensor technique, popular FCD sensors techniques include: automatic vehicle identification (AVI), license plate recognition(LPR), and transponders such as probe vehicles and electronic toll tags.
浮动车数据（FCD）主要是指ITS中不同位置的车辆移动数据，其中定制的探测器嵌入车辆中[32]。一些板载传感器为旅行路线选择和估计提供了有效和有效的信息。随着车辆传感器技术的发展，流行的FCD传感器技术包

括：自动车辆识别（AVI），车牌识别（LPR）以及诸如探测车辆和电子收费标签的转发器。

Wide area data refers to the wide area traffic flow data that is collected by diverse sensor tracking techniques such as photogrammetric processing, sound recording, video processing, and space-based radar.
广域数据是指通过各种传感器跟踪技术（如摄影测量处理，录音，视频处理和空基雷达）收集的广域流量数据。

E. Big Data From CAV and VANET
E.来自CAV和VANET的大数据

Connected and autonomous vehicles (CAV) are new technologies in ITS area that combines radical changes of vehicles design and their interactions with the road infrastructure. Connected and autonomous vehicles incorporate a range of different technologies, facilitating the safe, efficient movement of people and goods. CAV enabled traffic system has demonstrated great potential to mitigate congestion, reduce travel delay, and enhance safety performance [33], [34]. CAVs can generate big amount of environmentally relevant realtime transportation data, such as coordinate, speed, acceleration, safety data [33]. Using latest network technologies such as Software Defined Networking, data can be obtained more efficiently [35] These data can be used to create actionable information to support and facilitate green transportation choices, and apply to the real-time adaptive signal control [36], [37].
联网和自动驾驶车辆（CAV）是ITS领域的新技术，它结合了车辆设计的根本变化以及它们与道路基础设施的相互作用。联网和自动驾驶汽车采用了一系列不同的技术，促进了人员和货物的安全，有效运输。启用CAV的交通系统已显示出缓解拥堵，减少行程延误和提高安全绩效的巨大潜力[33]，[34]。CAV可以生成大量与环境相关的实时运输数据，例如坐标，速度，加速度，安全数据[33]。使用软件定义网络等最新网络技术，可以更有效地获取数据[35]这些数据可用于创建可操作的信息，以支持和促进绿色交通选择，并应用于实时自适应信号控制[36]， [37]。

Vehicle Ad Hoc Network (VANET) is a kind of mobile ad hoc network that uses vehicles and infrastructure elements as nodes to increase the coverage area and the communication capabilities. As an important part of ITS, VANET generates large amounts of data [38]. Data preparation and real-time results are challenging tasks for large-scale analysis. Using Big Data analytics, we can address most of data related VANET challenges [39], such as data filting [40], congestion and accidents alerting [41], and Traffic Flow prediction [42].
车载Ad Hoc网络（VANET）是一种移动ad hoc网络，它使用车辆和基础设施元素作为节点来增加覆盖区域和通信能力。作为ITS的重要组成部分，VANET会产生大量数据[38]。数据准备和实时结果是大规模分析的挑战性任务。使用大数据分析，我们可以解决大多数与VANET相关的数据挑战[39]，例如数据填充[40]，拥塞和事故警报[41]以及交通流量预测[42]。

F. Big Data From Passive Collection
F.被动收集的大数据

Compared with the actively collected data in transportation research, the rapid development of mobile technologies have enabled the collection of a massive amount of passive data. Passive data refers to those data not collected through active collection. It is generated for purposes that are not intended but can be potentially used for research [43], [44]. Chen et al. [45] and Zeyu et al. [46] propose to combine passive Big Data such as mobile phone data, internet access data and active data to study human mobility, travel behavior, and transportation planning. In [47], contextual information such as current time, cell phone ID, user identity are used for predicting the stay time of mobile users.

与运输研究中积极收集的数据相比，移动技术的快速发展使得能够收集大量的被动数据。被动数据是指未通过主动收集收集的数据。它的生成目的不是有意的，但可以用于研究[43]，[44]。陈等人。 [45]和Zeyu等人。 [46]建议结合被动大数据，如手机数据，互联网访问数据和活动数据，以研究人员流动性，旅行行为和交通规划。在[47]中，诸如当前时间，手机ID，用户身份的上下文信息用于预测移动用户的停留时间。

Social media data is the most popular passive data, and it refers to applications or websites where people interact with each other to create, share, and exchange information and ideas. Social media networks such as Linkedin, Facebook and Twitter have been developed rapidly recently. They have become relevant interests of transportation professionals as they provide information flows between providers and consumers in real time [48]. Though data collected via social media networks is generally unstructured and requires complicated processing, it provides significant transportation information when attitudes are expressed in different kind of transportation, and responses to travel disruptions are found in social media [49]–[52].
社交媒体数据是最受欢迎的被动数据，它指的是人们互相交互以创建，共享和交换信息和想法的应用程序或网站。最近，Linkedin，Facebook和Twitter等社交媒体网络迅速发展起来。它们已成为运输专业人员的相关利益，因为它们实时地在供应商和消费者之间提供信息流[48]。虽然通过社交媒体网络收集的数据通常是非结构化的并且需要复杂的处理，但是当在不同类型的交通中表达态度时，它提供了重要的交通信息，并且在社交媒体中发现了对旅行中断的响应[49] - [52]。

G. Big Data From Other Sources
G.来自其他来源的大数据

There are some sources of data that cannot be classified into the above categories. For example, real-time infrastructure state is considered as an important source of data [53]. The best known example is the smart grid [54], which will allow us to collect daily electricity consumption information for electric vehicles and train traction in urban rail transportation system. Another important data source is the data from dedicated test in ITS. For example, in our previous work, we carry out field tests in a real train ground communication system in urban rail transportation Communication Based Train Control (CBTC) system [55], [56]. A large amount of channel gain data is obtained from the field test. The data is processed to model the stochastic characteristic of channel state, and the model is used to optimize the CBTC system performance.
有些数据来源无法分类为上述类别。例如，实时基础设施状态被认为是重要的数据来源[53]。最着名的例子是智能电网[54]，它将使我们能够收集电动汽车的日常用电信息和城市轨道交通系统中的列车牵引力。另一个重要的数据来源是ITS中专用测试的数据。例如，在我们之前的工作中，我们在城市轨道交通基于通信的列车控制（CBTC）系统[55]，[56]的真实列车地面通信系统中进行现场测试。从现场测试中获得大量的信道增益数据。处理数据以模拟信道状态的随机特性，并且该模型用于优化CBTC系统性能。

# IV. BIG DATA ANALYTICS METHODS ITS
# IV。大数据分析方法

Machine learning is most popular modelling and analytics theory in Big Data ecosystems, which makes it easy to derive patterns and models from large amount of data. In ITS areas, machine learning theory has also be widely used to conduct data analytic. Depending on the completeness of data set that is available for learning, Machine learning models can be categorized into supervised, unsupervised and reinforcement learning algorithms. With the recent rapid

development of Artificial Intelligence, the powerful deep learning models have also been adopted to ITS recently.
机器学习是大数据生态系统中最流行的建模和分析理论，可以轻松地从大量数据中获取模式和模型。在ITS领域，机器学习理论也被广泛用于进行数据分析。根据可用于学习的数据集的完整性，机器学习模型可以分为监督，无监督和强化学习算法。随着近年来人工智能的快速发展，最近ITS也采用了强大的深度学习模型。

A. Supervised Learning

A.监督学习

Labeled training data is used in supervised learning algorithms [57]. The models use input data and the target outputs (labels) to learn the function or map between them. Combined with the learned model and the input data, the unseen outputs can be predicted. Among all the supervised learning models, linear regression, decision trees, neural networks, and support vector machines, are the most frequently used in ITSs.
标记的训练数据用于监督学习算法[57]。模型使用输入数据和目标输出（标签）来学习它们之间的功能或映射。结合学习模型和输入数据，可以预测看不见的输出。在所有监督学习模型中，线性回归，决策树，神经网络和支持向量机是ITS中最常用的。

The function of regression is to explain the relationship between one dependent variable and one or more independent variables. Linear regression is the most commonly used supervised learning [58]. Linear regression is incredibly simple, robust, easy to interpret, and easy to code. Despite its simplicity, linear regression is particularly successful in various ITS scenarios, such as traffic flow prediction [59], traffic speed estimation [60], and transportation travel route evaluation [61]. A decision tree is a decision support tool that uses a tree-like graph to model decisions and their possible consequences [62]. Due to their portability, robustness and transparency, decision trees are widely used in various ITS scenarios, such as traffic accident detection [63], accident severity analysis [64] and travel mode choice [65].
回归函数用于解释一个因变量与一个或多个自变量之间的关系。线性回归是最常用的监督学习[58]。线性回归非常简单，强大，易于理解，易于编码。尽管简单，但线性回归在各种ITS情景中尤其成功，例如交通流量预测[59]，交通速度估算[60]和交通出行路线评估[61]。决策树是一种决策支持工具，它使用树状图来模拟决策及其可能的后果[62]。由于其可移植性，鲁棒性和透明性，决策树广泛用于各种ITS场景，例如交通事故检测[63]，事故严重性分析[64]和旅行模式选择[65]。

Artificial Neural network (ANN) is a popular example of flexible and robust supervised learning for both classification and regression [57]. With enough hidden layers of processing nodes and training data, ANN can learn any non-linear relations between input and target data. As a data modeling tool, it has also been adopted in ITS such as traffic flow prediction [66], travel time prediction [67], traffic accident detection [68] and remaining parking spaces forecasting [69]. Support vector machine (SVM) is another popular supervised learning algorithms that use labelled data for regression and classification. Among all the Big Data analytics model tools in ITS, SVMs have attracted great interests in research area. It has been successfully used in travel time prediction [70], bus arrival time prediction [71], and traffic accident detection [72].
人工神经网络（ANN）是分类和回归的灵活有力的监督学习的一个流行的例子[57]。通过处理节点和训练数据的足够隐藏层，ANN可以学习输入和目标数据之间的任何非线性关系。作为一种数据建模工具，它也被ITS采用，如交通流预测[66]，行程时间预测[67]，交通事故检测[68]和剩余停车位预测[69]。支持向量机（SVM）是另一种流行的监督学习算法，它使用标记数据进行回归和分类。在ITS中的所有大数据分析模型工具中，SVM引起了研究领域的极大兴趣。它已成功用于旅行时间预测[70]，公交车到达时间预测[71]和交通事故检测[72]。

A typical example of using supervised learning in ITS is introduced in [72], where SVM is used to predict traffic incidents. Given the training subset $\{(x1, y1), (x2, y2), \ldots, (xi, yi), \ldots\}$, where xi is the input of the training sample which consists of the values of the traffic flow parameters such as volume, speed, occupancy and so on, and yi is the class label of xi. With a kernel function $K(x, x')$, according to the SVM classifier theory, the support vector αi can be obtained as,

在[72]中引入了在ITS中使用监督学习的典型示例，其中SVM用于预测交通事故。给定训练子集$\{(x1, y1), (x2, y2), \ldots, (xi, yi), \ldots\}$，其中xi是训练样本的输入，其包括交通流量参数的值，例如音量，速度，占用率等，以及yi是xi的类标签。利用核函数$K(x, x')$，根据SVM分类器理论，支持向量αi可以得到，

$$\max_{\alpha_i} \; -\frac{1}{2}\sum_{i=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i, x_j) + \sum_{i=1}^{l} \alpha_i$$
$$s.t. \; \sum_{i=1}^{l} y_i \alpha_i = 0, \qquad (1)$$

Then, we get the decision function g(x ) to compute the label for the sample x as,

然后，我们得到决策函数g（x）来计算样本x的标签，

$$g(x) = sgn(\sum_{i=1}^{l} y_i \alpha_i^* K(x_i, x) + b). \qquad (2)$$

If x is an incident sample, $g(x) = 1$. Otherwise, we have $g(x) = -1$.

如果x是一个入射样本，$g(x) = 1$.否则，我们有$g(x) = -1$。

B. Unsupervised Learning

B.无监督学习

Unsupervised learning normally also referred as clustering focus on learning natural group from unlabeled multidimensional data [57]. K-means is the most popular unsupervised learning tool, and it has been widely adopted in highway transportation planning [73], and travel time prediction [74]. With a set of historical data, authors of [74] gives a classic example of using unsupervised learning to predict travel time. The procedures are as follows,

无监督学习通常也称为聚类，重点是从未标记的多维数据中学习自然群[57]。K-means是最受欢迎的无监督学习工具，它已被广泛应用于公路交通规划[73]和旅行时间预测[74]。通过一组历史数据，[74]的作者给出了使用无监督学习来预测旅行时间的经典例子。程序如下，

1. Compute the travel time frequency ε. It means the number of time that the travel time appears.

1.计算行程时间频率ε。它表示旅行时间出现的时间。

2. Define a tuple $\Gamma(\tau_i, \varepsilon_i, \nu_i)$ that contains distinct features, where τi is the travel time, εi is the travel time frequency, and νi is the travel velocity.

2.定义一个元组$\Gamma(\tau_i, \varepsilon_i, \nu_i)$包含不同的特征，其中τi是行程时间，εi是行程时间频率，νi是行进速度。

3. Find the greatest value in the data based on the travel time frequency. A tuple $\Gamma(\tau_p, \varepsilon_p, \nu_p)$ is chosen as a centroid of Cluster 1, where $\varepsilon_p$ is the maximum travel time frequency, $\tau_p$ is the corresponding maximum travel time associated with $\varepsilon_p$, and $\nu_p$ is the travel velocity associated with $\varepsilon_p$.

3.根据旅行时间频率找出数据中的最大值。选择元组$\Gamma(\tau_p, \varepsilon_p, \nu_p)$作为簇1的质心，其中εp是最大行程时间频率，τp是与εp相关的相应最大行程时间，νp是与之相关的行进速度εp。

4. Compare each tuple $\Gamma(\tau_i, \varepsilon_i, \nu_i)$ with the centroid $\Gamma(\tau_p, \varepsilon_p, \nu_p)$ of Cluster 1 by compute their distance. Choose the tuple $\Gamma(\tau_q, \varepsilon_q, \nu_q)$ with the maximum distance.

4.通过计算它们的距离，比较群集1的每个元组$\Gamma(\tau_i, \varepsilon_i, \nu_i)$与质心$\Gamma(\tau_p, \nu_p, \nu_p)$。选择元组$\Gamma(\tau_q, \varepsilon_q, \nu_q)$最大距离。

5. Build two clusters where the centroid of Cluster 1 is tuple $\Gamma(\tau_p, \varepsilon_p, \nu_p)$ and that of Cluster 2 is tuple $\Gamma(\tau_q, \varepsilon_q, \nu_q)$.

5.构建两个簇，其中簇1的质心是元组$\Gamma(\tau_p, \nu_p, \nu_p)$和簇2的质心是元组$\Gamma(\tau_q, \varepsilon_q, \nu_q)$。

6. Define the cluster memberships of all the tuples by assigning them to the nearest cluster centroid.

6.通过将所有元组分配给最近的集群质心来定义所有元组的集群成员资格。

7. Re-estimate the cluster centre using the arithmetic mean.

7.使用算术平均值重新估计群集中心。

8. Repeat step 6 and 7.

8.重复步骤6和7。

9. After complete preparation of clusters, desired predicted time is calculated separately for each cluster as, ςk time is calculated separately for each cluster as, $\varsigma_k = \sum_{j=1}^{N} \varepsilon_j * \tau_j / \sum_{j=1}^{N} \varepsilon_j$. Where ςk is the travel time obtained from kth cluster, N is the total number of tuple in the associated cluster, $\varepsilon_j$ is the travel time frequency of the j th tuple, and $\tau_j$ is the travel time of the j th tuple.

9.在完成簇的准备之后，针对每个簇分别计算期望的预测时间，因为ηk time is calculated separately for each cluster as, $\varsigma_k = \sum_{j=1}^{N} \varepsilon_j * \tau_j / \sum_{j=1}^{N} \varepsilon_j$。其中ζk是从第k个聚类获得的旅行时间，N是相关聚类中元组的总数，εj是第j个元组的旅行时间频率，τj是第j个元组的旅行时间。

10. The final predicted approximate travel time is obtained by computing the arithmetic mean of ς1 and ς2.

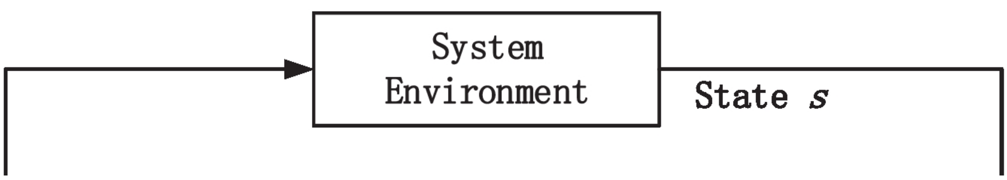
10.最终预测的近似旅行时间是通过计算ς1和ζ2的算术平均值得到的。

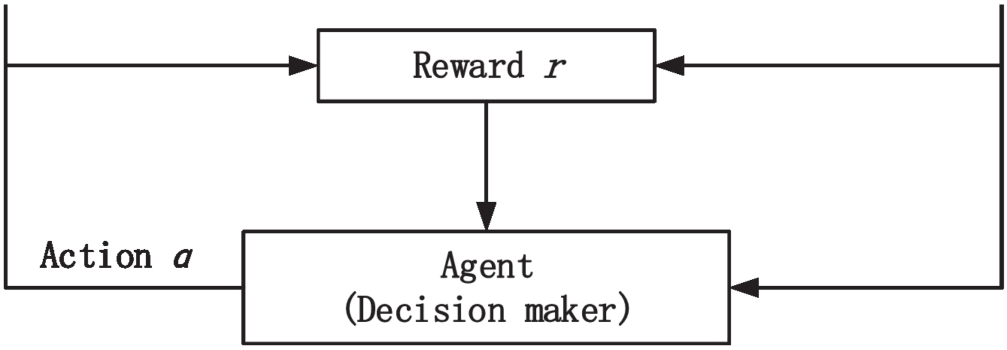
C. Reinforcement Learning

C.强化学习

Different from supervised and unsupervised learning, as shown in Fig. 2, the aim of the reinforcement learning is to minimize the long term cost through exploration and Fig. 2. Reinforcement learning.

与有监督和无监督学习不同，如图2所示，强化学习的目的是通过探索最小化长期成本，图2.强化学习。






learn the optimal policy by interacting with the experimental data [57]. Reinforcement learning is highly relevant to control and optimization theory, and it has been proved to be quite feasible in traffic signal control in ITS [75]–[78].
通过与实验数据相互作用来学习最优政策[57]。强化学习与控制和优化理论高度相关，并且已经证明它在ITS中的交通信号控制中是非常可行的[75] - [78]。

Using reinforcement learning in ITS requires a formulation of the ITS control and optimization problem in the language of reinforcement learning, specifically, defining a state space S, an action space A and a reward R. One classic example of using reinforcement learning in ITS traffic signal control is formulated in [76]. The state of traffic at an intersection with n lanes is formally defined as the discrete traffic state encoding (DTSE). For each lane approaching the intersection, the DTSE discretizes a length l of the lane segment, beginning at the stop line, into cells of length c. The selection of c will change the behavior of system. The DTSE is composed of three vectors. The first vector B represents the presence of a vehicle or not in the cell. The second vector R represents the speed of the vehicle, and the third vector P is the current traffic signal phase (i.e., the most recent action selected). Thus, the system states can be defined as, S ∈ ( B R) lc n P.
在ITS中使用强化学习需要在强化学习的语言中制定ITS控制和优化问题，具体地说，定义状态空间S，动作空间A和奖励R.在[76]中制定了在ITS交通信号控制中使用强化学习的一个典型例子。在具有n个通道的交叉点处的交通状态被正式定义为离散交通状态编码（DTSE）。对于接近交叉路口的每个车道，DTSE将从停车线开始的车道段的长度l离散为长度为c的单元。选择c将改变系统的行为。 DTSE由三个向量组成。第一个向量B表示在细胞中是否存在载体。第二矢量R表示车辆的速度，第三矢量P是当前的交通信号相位（即，所选择的最近的动作）。因此，系统状态可以定义为，S ∈ ( B R） lc n P.

After the agent has observed the state of the environment, it must choose one action from the set of all available actions. The possible actions are North-South Green (a1), East-West Green (a2)), North-South Advance Left Green (a3)), East-West Advance Left Green (a4)). The set of all possible actions A is defined as A = { a1, a2, a3, a4}. At time t, the agent chooses an action a(t ), where a(t ) ∈ A.
在代理观察到环境状态后，它必须从所有可用操作集中选择一个操作。可能的行动是南北绿（a1），东西绿（a2）），南北推进左绿（a3）），东西推进左绿（a4））。所有可能的动作A的集合被定义为A = { a1，a2，a3，a4}。在时间t，代理选择动作a（t），其中a（t ) ∈ A.

After the agent has observed the state of the environment st , it performs an action a(t ), and receives the reward. The reward rt +1 is a consequence of performing a selected action from a specific state. In this formulation, the reward is defined as change in cumulative vehicle delay between actions.
在代理观察到环境st的状态之后，它执行动作a（t），并接收奖励。奖励rt + 1是从特定状态执行选定动作的结果。在此公式中，奖励被定义为行动之间累积车辆延迟的变化。

The reinforcement learning algorithm used in this formulation is Q-Learning [57], which is used to develop an optimal

action-selection policy. The optimal policy is achieved by using the convolutional neural network to approximate the action-value function. The action-value function Q(st , at ) maps states to action utilities (i.e., what is the value of each action from a given state). The basis of Q-learning is the value iteration update defined as,
该公式中使用的强化学习算法是Q-Learning [57]，用于制定最优的动作选择策略。通过使用卷积神经网络来近似动作 - 值函数来实现最优策略。动作值函数Q（st，at）将状态映射到动作实用程序（即，来自给定状态的每个动作的值是多少）。Q学习的基础是价值迭代更新定义为，

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha(r_{t+1} + \gamma \max_A Q(s_{t+1}, a_t) - Q(s_t, a_t)). \quad (3)$$

Where the learning rate α controls the degree to which new action-value estimates are weighted against old estimates and the discount factor γ determines how immediate rewards are weighted against future rewards. After the action-value function has been sufficiently learned, the optimal policy can be determined by selecting the action with the highest value.
学习率α控制新行动 - 价值估计与旧估计加权的程度，折扣因子γ确定如何立即对未来奖励加权。在充分学习了动作值功能之后，可以通过选择具有最高值的动作来确定最优策略。

D. Deep Learning
D.深度学习

Deep learning models exploit much more system features and complex architecture than traditional Artificial Neural Network, and can achieve better performance than traditional machine learning models. They have been widely applied in ITSs. For example, a deep Restricted Boltzmann Machine and Recurrent Neural Network architecture is utilized to model and predict traffic congestion evolution based on GPS data from taxi [79]. Using deep neural networks, fault diagnoses on bogies with Big Data is carried out in [80]. Chen [81] carry out the vehicle detection task using the rich feature of convolutional neural network(CNN) learned from ImageNet dataset. Duan et al. [82] use stacked auto-encoders for traffic data imputation. In traffic flow area, deep learning model has become a popular tool to predict traffic flow density [83]–[86]. Literature [85] gives a typical deep learning based approach to do the traffic flow prediction. Stacked autoencoders (SAEs) are used to learn generic traffic flow features. Considering SAEs with K layers, the first layer is trained as an autoencoder, with the training set as inputs. After obtaining the first hidden layer, the output of the j th hidden layer is used as the input of the (k + 1)th hidden layer. In this way, multiple autoencoders can be stacked hierarchically. To use the SAE network for traffic flow prediction, a logistic regression layer is added on top of the network for supervised traffic flow prediction. The whole deep architecture model is shown in Fig.3.
深度学习模型比传统的人工神经网络利用更多的系统特征和复杂的架构，并且可以实现比传统机器学习模型更好的性能。它们已被广泛应用于ITSs。例如，利用深度限制玻尔兹曼机器和递归神经网络架构来模拟和预测基于出租车的GPS数据的交通拥堵演变[79]。使用深度神经网络，在大数据的转向架上进行故障诊断在[80]中进行。Chen [81]利用从ImageNet数据集中学习的卷积神经网络（CNN）的丰富特征来执行车辆检测任务。段等人。 [82]使用堆叠式自动编码器进行交通数据插补。在交通流域，深度学习模型已成为预测交通流密度的流行工具[83] - [86]。文献[85]给出了一种典型的基于深度学习的方法来进行交通流量预测。堆叠自动编码器（SAE）用于学习通用流量传输功能。考虑具有K层的SAE，第一层被训练为自动编码器，训练集作为输入。在获得第一隐藏层之后，第j个隐藏层的输出被用作第（k + 1）个隐藏层的输入。以这种方式，可以分层堆叠多个自动编码器。为了

使用SAE网络进行交通流量预测，在网络顶部添加了逻辑回归层，用于监督交通流量预测。整个深层架构模型如图3所示。

The data collected from all freeways are used as the input. Considering the temporal relationship of traffic flow, the traffic flow data at previous time intervals, i.e., χt −1 , χ t −2,. . , χt −l are used to predict the traffic flow at time interval t. The proposed model accounts for the spatia land temporal correlations of traffic flow inherently.

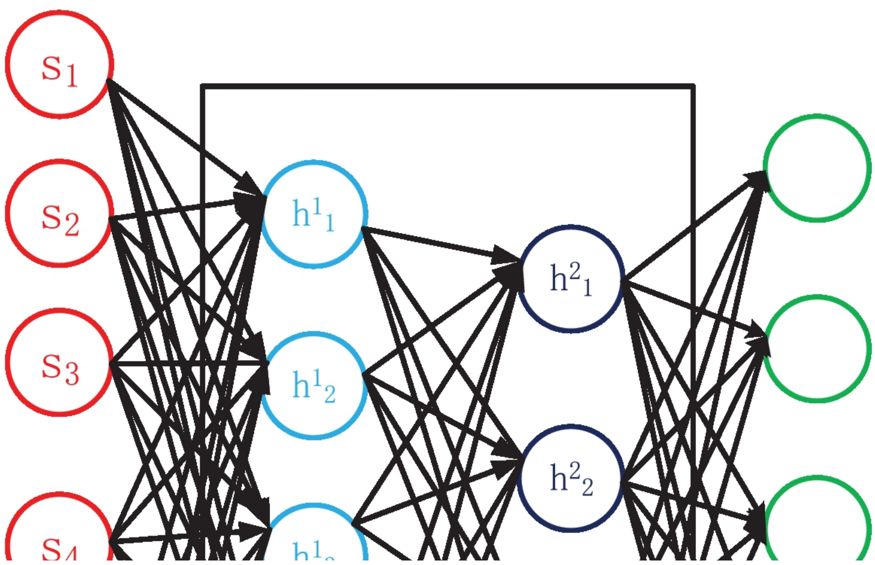
从所有高速公路收集的数据用作输入。考虑到交通流量的时间关系，交通流量数据在先前的时间间隔，即χt-1 , χ t -2。 ，χt-1用于预测时间间隔t的交通流量。所提出的模型固有地解释了交通流的时空相关性。

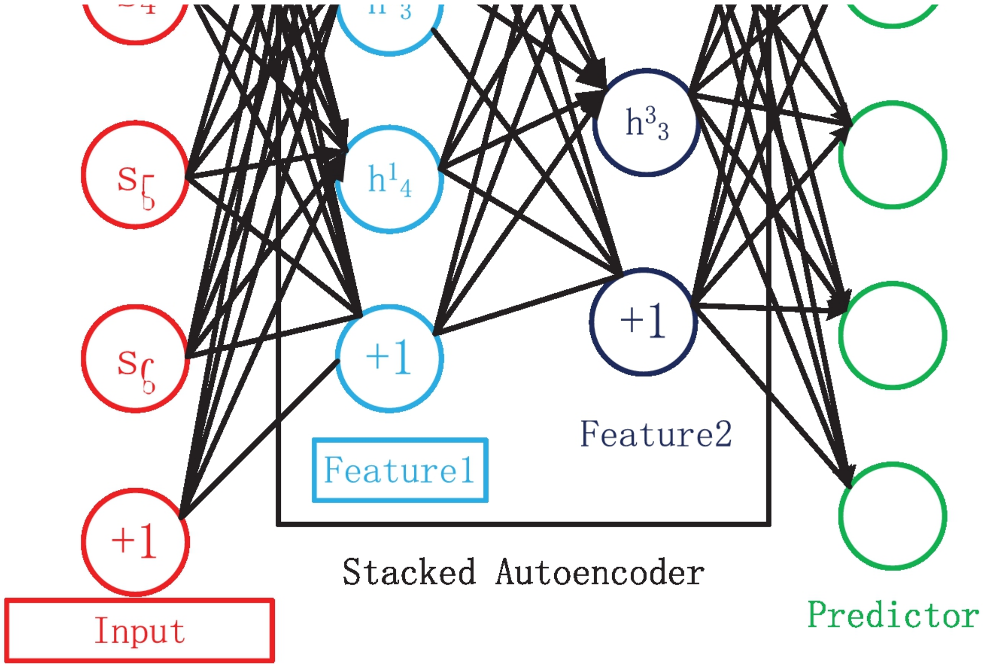
E. Ontology Based Methods

E.基于本体的方法

An ontology is a formal naming and definition of the types, properties, and interrelationships of the entities that really exist in a particular domain of discourse. Ontology based methods can accurately describe data semantics and infer implicit data semantic relations. Compared with the traditional data extraction from the bottom up, ontology data integration has a topdown feature and uses ontology modeling to share semantic views of data and map heterogeneous data from different data sources to minimize or even eliminate the ambiguous understanding of shared data. Ontology based method has been widely applied in ITSs. For example, Zhai et al. [87] propose an information retrieval system for ITS based on a fuzzy ontology framework. This framework includes three parts: concepts, properties of concepts and values of properties, and it focuses on information about traffic accidents. Fernandez and Ito [88] propose a driver behavior model based in ontology for intelligent transportation system. The driver behavior ontology has the knowledge related to driver characteristics, perception and cognitive state to perform different driving Fig. 3. Deep architecture model for traffic flow prediction.

本体论是对特定话语领域中存在的实体的类型，属性和相互关系的正式命名和定义。基于本体的方法可以准确地描述数据语义并推断隐式数据语义关系。与自下而上的传统数据提取相比，本体数据集成具有自上而下的特征，并使用本体建模来共享数据的语义视图并映射来自不同数据源的异构数据，以最小化甚至消除对共享数据的模糊理解。基于Ontology的方法已在ITS中得到广泛应用。例如，Zhai等人。 [87]提出了一种基于模糊本体框架的ITS信息检索系统。该框架包括三个部分：概念，概念属性和属性值，并侧重于有关交通事故的信息。Fernandez和Ito [88]提出了一种基于本体的智能交通系统驾驶员行为模型。驾驶员行为本体具有与驾驶员特征，感知和认知状态相关的知识以执行不同的驾驶（图3）。交通流预测的深度架构模型。






tasks. It can be used to predict traffic accidents and optimize the road congestion. Fernandez and Ito [89] propose to use the ontology to manage the sensor information in intelligent transportation systems and convert the sensor data into semantic data. The system performs the automatic traffic light settings can use the data to predict and avoid traffic accidents. Gregor et al. [90] propose a systematic methodology to create ontology in ITS domain. This ontology will serve as the basis of semantic information to a semantic service that allows the connection of new equipment to an urban network. Zhao et al. [91] introduce an ontology-based Knowledge Base, which contains maps and traffic regulations. By accessing to the Knowledge Base, the intelligent vehicles can be aware of over speed situations and make decisions at intersections in comply with traffic regulations. Chen et al. [92] depict an ontology-based approach for safety management in Cooperative ITS (C-ITS), primarily in an automotive context. It provides the support for ontology driven ITS development and its formal information model. Yang and Wang [93] take advantages of the semantic completeness of the ontology to build urban traffic ontology model, which resolve the problems as ontology mergence and equivalence verification in semantic fusion of traffic information integration. The model can increase the function of semantic fusion, and reduce the amount of data integration of urban traffic information as well enhance the efficiency and integrity of traffic information query.

任务。它可用于预测交通事故并优化道路拥堵。Fernandez和Ito [89]建议使用本体来管理智能交通系统中的传感器信息，并将传感器数据转换为语义数据。系统执行自动交通灯设置可以使用数据来预测和避免交通事故。格雷戈尔等人。 [90]提出了一种在ITS领域创建本体的系统方法。该本体将作为语义服务的语义信息的基础，该语义服务允许将新设备连接到城市网络。赵等人。 [91]介绍了一个基于本体的知识库，其中包含地图和交通法规。通过访问知识库，智能车辆可以了解超速情况，并在交叉路口做出符合交通法规的决策。陈等人。 [92]描述了一种基于本体的协同ITS（C-ITS）安全管理方法，主要用于汽车领域。它为本体驱动的ITS开发及其正式的信息模型提供支持。Yang和Wang [93]利用本体的语义完备性来构建城市交通本体模型，解决了交通信息集成语义融合中的本体合并和等价验证等问题。该模型可以增加语义融合的功能，减少城市交通信息的数据整合量，提高交通信息查询的效率和完整性。

## V. BIG DATA APPLICATIONS IN ITS
## V.大数据应用

Big Data provides technical supports for the development and applications of ITS. By efficient, accurate and timely data collection, analyzing and processing in road and rail transportation system, the Big Data applications can provide the public with convenient and high efficient transportation. In order to identify problems, improving ITS efficiency, reducing costs and deriving valuable insights, Big Data applications in ITS can be divided into the following six categories.

大数据为ITS的开发和应用提供技术支持。通过高效，准确，及时的公路和铁路运输系统数据采集，分析和处理，大数据应用可以为公众提供方便，高效的运输。为了识别问题，提高ITS效率，降低成本并获得有价值的见解，ITS中的大数据应用可分为以下六类。

A. Road Traffic Accidents Analysis

A.道路交通事故分析

Evidence shows that in the world around 1.2 million people are killed and 50 million injured from traffic accidents every year [94]. Accurate traffic accident data analysis results can provide traffic department with important information to make policies so as to prevent accidents.

有证据表明，全世界每年约有120万人死于交通事故，造成5000万人受伤[94]。准确的交通事故数据分析结果可以为交通部门提供重要信息，以制定政策以防止事故发生。

Many studies focused on using Big Data analytics in traffic accidents analysis. Using measured traffic flow data, Golob and Recker [95] study the relationships among weather, lighting conditions, traffic flow, and urban freeway accidents, with a multivariate statistical model. In [10], Bayesian inference and Random forest are adopted in a real-time crash prediction model to reduce crash risks. Xiong et al. [96] introduce classification and regression trees (CART), logistic regression and multivariate adaptive regression splines (MARS) to perform analytical operation on motor vehicle accident injury data. Lee and Mannering [97] present a method which uses zero-inflated count models and nested logit models to analyze run-off-roadway accident frequency and severity on a 96.6km section of highway in Washington State. The results show that some measures can be taken to reduce run-off-roadway accident frequencies. Karlaftis and Golias [98] apply a rigorous non-parametric statistical methodology which is hierarchical tree-based regression (HTBR) to analyze the influences of terrain and traffic characteristics on accident rates of rural roads. The methodology can also be used to predict the accident rates of highway. Chang, et. analyze the relationship between highway geometric variables and traffic accidents by using a negative binomial regression model and a classification and regression tree model. The parameters come from the 2001-2002 accident data of National Freeway 1 in Taiwan [99]. Bédard et al. [100] determine the respective effect of driver, crash, and vehicle characteristics to the fatality risk of drivers by using a multivariate logistic regression algorithm, the results indicate that increasing seat belt use, reducing vehicle speed, and decreasing the number and severity of driver-side impacts could prevent traffic accidents.

许多研究都侧重于在交通事故分析中使用大数据分析。使用测量的交通流量数据，Golob和Recker [95]利用多元统计模型研究了天气，光照条件，交通流量和城市高速公路事故之间的关系。在[10]中，在实时碰撞预测模型中采用贝叶斯推理和随机森林来降低碰撞风险。熊等人[96]引入分类和回归树（CART），逻辑回归和多元自适应回归样条（MARS）来执行机动车事故伤害数据的分析操作。Lee和Mannering [97]提出了一种方法，该方法使用零入射计数模型和嵌套logit模型来分析华盛顿州96.6km高速公路上的径流事故频率和严重程度。结果表明，可以采取一些措施来减少径流事故频率。Karlaftis和Golias [98]采用严格的非参数统计方法，即基于分层树的回归（HTBR）来分析地形和交通特征对农村道路事故率的影响。该方法还可用于预测高速公路的事故率。张，等。

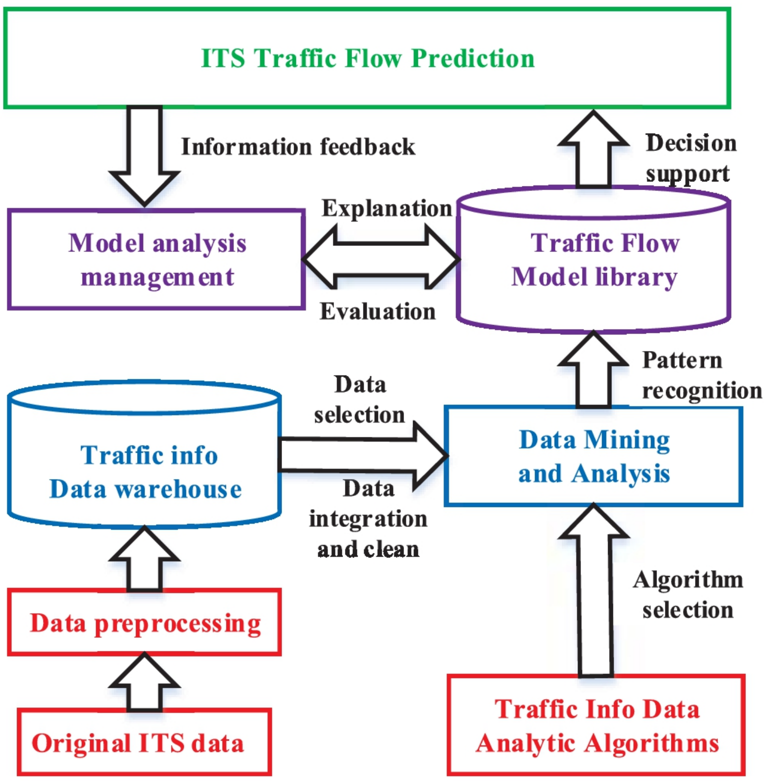
通过使用负二项回归模型和分类和回归树模型分析公路几何变量与交通事故之间的关系。这些参数来自台湾国家高速公路1号的2001 - 2002年事故数据[99]。Bédard等。 [100]通过使用多变量逻辑回归算法确定驾驶员，碰撞和车辆特征对驾驶员死亡风险的相应影响，结果表明增加安全带使用，降低车速，减少驾驶员的数量和严重程度侧面撞击可以防止交通事故。

B. Road Traffic Flow Prediction

B.道路交通流量预测

Timely and accurate traffic flow information is critical for transportation management. Big Data analytics in ITS has an advantage in traffic flow prediction [101]–[103]. According to [9], a classic road traffic flow prediction model using Big Data analytics is shown in Fig 4. The original ITS data is first preprocessed to get the effective data set. Using selected data mining or analysis method, traffic flow model is established with the preprocess data. The traffic flow model gives decision supports to traffic management department and get feedback from real traffic flows to calibrate the model.
及时准确的交通信息对于运输管理至关重要。ITS中的大数据分析在交通流量预测方面具有优势[101] - [103]。根据[9]，使用大数据分析的经典道路交通流预测模型如图4所示。首先对原始ITS数据进行预处理以获得有效数据集。使用选定的数据挖掘或分析方法，使用预处理数据建立交通流模型。交通流模型为交通管理部门提供决策支持，并从实际交通流中获得反馈，以校准模型。

Many scholars have studied traffic flow prediction using Big Data analytic. Lv et al. [85] propose a deep learning based Fig. 4. A typical traffic flow prediction model.
许多学者使用大数据分析研究了交通流量预测。Lv等人。 [85]基于图4提出深度学习。典型的交通流预测模型。




traffic flow prediction method which use the greedy layerwise unsupervised learning algorithm. Stacked auto encoder (SAE) model is used to learn generic traffic flow features. The results show that the deep learning based model has superior performance for traffic flow prediction. Liu et al. [104] analyze multidimensional parameters and the traffic flow

prediction models is developed from different dimensions based on SVMs. Dong et al. [105] propose a pre-selection space time model to estimate the traffic flow at locations with little data detectors. Canaud et al. [106] present a probability hypothesis density filtering based model for real-time traffic flow prediction. Pan et al. [107] put forward a modified stochastic cell transmission model to support short-term traffic flow prediction. Antoniou et al. [108] propose an approach for local traffic flow state estimation and prediction based on data-driven computational approaches. Using the seemingly unrelated time-series equation (SUTSE), Ghosh et al. [109] present a new multivariate structural time-series (MST) model to predict traffic flow. The SUTSE model can respectively track the change of each traffic flows and their components as time goes by, and the results show it has a superior prediction accuracy. Xu et al. [110] propose a novel online algorithm which is a context-aware adaptive traffic prediction algorithm. The algorithm can learn from the current traffic condition and use the historical traffic data to predict the future traffic flow. The experiments indicate that this algorithm do better than the current solutions. Lu et al. [111] build a traffic flow state clustering model which adopts the simulated annealing genetic algorithm using fuzzy c-means (SAGA-FCM). This model is based on traffic speed data and occupancy data which comprehensively considers the temporal, spatial, and historical correlations of traffic flow Big Data.
利用贪婪分层无监督学习算法的交通流预测方法。堆叠式自动编码器（SAE）模型用于学习通用流量传输功能。结果表明，基于深度学习的模型在交通流预测中具有优越的性能。刘等人。 [104]分析多维参数，交通流预测模型是基于SVM从不同维度开发的。董等人。 [105]提出了一种预选空间时间模型，用于估算数据检测器较少的位置的交通流量。Canaud等人。 [106]提出了一种基于概率假设密度滤波的模型，用于实时交通流量预测。潘等人。 [107]提出了一种改进的随机细胞传递模型，以支持短期交通流量预测。Antoniou等。 [108]提出了一种基于数据驱动计算方法的局部流量状态估计和预测方法。使用看似无关的时间序列方程（SUTSE），Ghosh等。 [109]提出了一种新的多变量结构时间序列（MST）模型来预测交通流量。随着时间的推移，SUTSE模型可以分别跟踪每个交通流量及其组成部分的变化，结果表明它具有更高的预测精度。徐等人。 [110]提出了一种新颖的在线算法，它是一种上下文感知自适应交通预测算法。该算法可以从当前的交通状况中学习，并使用历史交通数据来预测未来的交通流量。实验表明，该算法比目前的解决方案做得更好。Lu等人。 [111]建立一个交通流状态聚类模型，采用模糊退火遗传算法，采用模糊c-均值（SAGA-FCM）。该模型基于交通速度数据和占用数据，全面考虑交通流量大数据的时间，空间和历史相关性。

With the recent rapid development of AI technology, deep learning methods have been widely applied to predict traffic flow. Huang et al. [84] introduce deep belief network into transportation system. Ma et al. [79] combined deep restricted Boltzmann machines (RBM) with RNN and formed a RBM-RNN model that inherits the advantages of both RBM and RNN. They also [86] use LSTM to predict traffic and demonstrate that LSTM achieve better performance compared with traditional neural networks in both stability and accuracy regarding traffic speed prediction by using loop detector data collected in the Beijing road network. Lv et al. [85] propose a novel deep-learning- based traffic prediction model that considered spatiotemporal relations, and employed stack autoencoder (SAE) to extract traffic features.
随着近来AI技术的快速发展，深度学习方法已被广泛应用于预测交通流。黄等人。 [84]将深厚的信念网络引入交通系统。Ma等人。 [79]将深度限制玻尔兹曼机（RBM）与RNN相结合，形成了RBM-RNN模型，该模型继承了RBM和RNN的优点。他们还[86]使用LSTM预测交通，并证明LSTM与传统神经网络相比，通过使用北京公路网收集的环路探测器数据，在稳定性和准确性方面实现了更好的性能。Lv等人。 [85]提出了一种新的基于深度学习的交通预测模型，该模型考虑了时空关系，并采用堆栈自动编码器（SAE）来提取交通特征。

C. Public Transportation Services Planning

C.公共交通服务规划

Public transportation Big Data analytics can help to understand transportation riders journey patterns across the transportation network. The riders journey patterns can be used to inform decisions to transportation operators about the services planning.

公共交通大数据分析有助于了解交通网络中的交通乘客旅程模式。乘客行程模式可用于向运输运营商通报有关服务计划的决策。

With heterogeneous sources of traffic measurements data, Lu et al. [112] present a path flow based nonlinear optimization model to estimate dynamic OD demand that does not need explicit dynamic link information. Using triangulated mobile phone records of millions of anonymous users, authors of [113] present a method to predict average daily OD trips. The applicability of the proposed model is verified by the spatial and temporal distributions of trips get from local and national surveys. Using complete daily set of smart card data from London Metro and iBus vehicle location system, Gordon [114] derives the boarding and alighting times of every passenger, and transfer information is derived from passenger trips belong to different public transportation modes. The full journey matrices are established from the data, and are validated by traditional O-D matrices. The approach is efficient enough to be performed daily and provide the transportation operators travel behavior of their services. The Big Data analytics results in these works can help the emerging intelligent traffic management applications generate proactive, coordinated traffic information provision. Tao [115] investigate the temporal and spatial dynamics of Bus Rapid Transit (BRT) trips against non-BRT trips during five typical calendar events. The smart card data is first pre-processed to build OD flow matrices and bus trip route for BRT and non-BRT trips respectively. Service management department can identify important implications for evidence-based BRT policies. In [116], operational Big Data from Automated Fare Collection (AFC) systems is used for transportation planning management in Istanbul, Turkey. Works in MIT [117] shows the potential value of London AFC data in rail transportation planning and operations. The applications developed in their work provide rail transportation operators and planner an easy-to-update management tool that evaluates rail service in several aspects at near real-time. Toole et al. [118] use mobile phone data from open source data repositories to implement a travel demand model. Routable road networks, validated OD matrices and trip tables can be extract from the Call Data Record (CDR) data with the model. Their work serve as universal guide to help the transportation operators perform public transportation planning.

对于交通测量数据的异构来源，Lu等人。 [112]提出了一种基于路径流的非线性优化模型，用于估计不需要显式动态链接信息的动态OD需求。使用数百万匿名用户的三角手机记录，[113]的作者提出了一种预测平均每日OD旅行的方法。所提出的模型的适用性通过来自地方和国家调查的旅行的空间和时间分布来验证。Gordon [114]使用来自伦敦地铁和iBus车辆定位系统的完整日常智能卡数据，得出每位乘客的登机和下车时间，并且转乘信息来自属于不同公共交通方式的旅客出行。完整的旅程矩阵是根据数据建立的，并由传统的O-D矩阵验证。该方法足以每天执行，并为运输运营商提供其服务的旅行行为。这些工作中的大数据分析结果可以帮助新兴的智能交通管理应用程序生成主动，协调的交通信息提供。Tao [115]研究了五次典型日历事件中快速公交（BRT）旅行对非BRT旅行的时空动态。首先对智能卡数据进行预处理，以分别为BRT和非BRT行程构建OD流量矩阵和总线跳闸路径。服务管理部门可以识别基于证据的BRT政策的重要含义。在[116]中，自动收费系统（AFC）系统的运营大数据用于土耳其伊斯坦布尔的交通规划管理。麻省理工学院的工作[117]显示了伦敦AFC数据在铁路运输规划和运营中的潜在价值。在他们的工作中开发的应用程序为铁路运输运营商和规划人员提供了一种易于更新的管理工具，可以近乎实时地评估几个方面的铁路服务。Toole等。 [118]使用来自开源数据存储库的移动电话数据来实现旅行需求

模型。可以从模型中的呼叫数据记录（CDR）数据中提取可路由的道路网络，经验证的OD矩阵和行程表。他们的工作是帮助运输经营者执行公共交通规划的通用指南。

D. Personal Travel Route Planning

D.个人旅行路线规划

The transportation Apps start with great vision. Report suggests that only telling passengers the arriving time of the next bus could make them more satisfied with the bus service [119]. Based on the data from smart phones and vehicle GPS data, some transportation APPs provide riders with real time traffic information [120], others provide most suitable driving routes with minimum travel time [121]. Combined with public transportation data with information from users through their smart phones, transportation APPs can even provide riders with real time public transportation journey planning [122]. Fully integrated Apps even let people plan trips that move from trains to buses and private cars or bicycles at the ends [123].
运输应用程序始于良好的愿景。报告建议只告诉乘客下一班车的到达时间可以使他们更满意公交服务[119]。根据智能手机和车辆GPS数据的数据，一些运输APP为乘客提供实时交通信息[120]，其他提供最合适的行车路线，最短的旅行时间[121]。通过智能手机将公共交通数据与用户信息相结合，交通APP甚至可以为乘客提供实时的公共交通旅程规划[122]。完全集成的应用程序甚至可以让人们计划从火车到公共汽车和私家车或自行车的旅程[123]。

Big Data analytics in these transportation APPs generates huge economic benefits by reducing travel time, traffic congestion, pollution, and greenhouse-gas emissions. For example, opening up Transportation for London (TfL) data has been valued at 15-58 million pounds per year and has resulted in over 200 travel Apps being developed by private companies [124].
这些运输APP中的大数据分析通过减少旅行时间，交通拥堵，污染和温室气体排放产生了巨大的经济效益。例如，开放伦敦交通局（TfL）数据的价值每年为15-58百万英镑，并且私人公司开发了200多个旅行应用程序[124]。

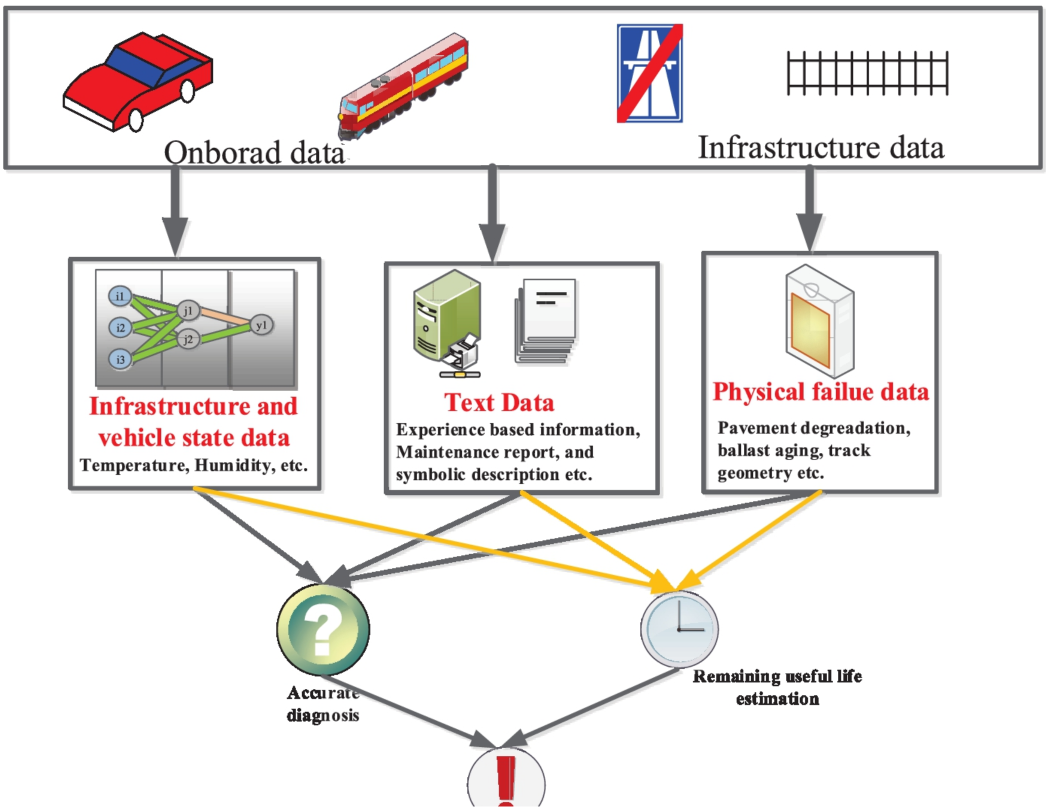
E. Rail Transportation Management and Control

E.铁路运输管理和控制

Rail transportation systems have been transformed with advanced IT technology. They are the main beneficiary of Big Data analytics. This is because that Rail transportation systems are generally closed systems that carry out sophisticated processing of large volumes of data, such as real time train speed and position, train departure and arrival time of a certain station, and passenger OD information. Big Data analytics can make the rail transportation operators be better at train control and improve the rail transportation system operation efficiency. In industry, Big Data analytics is starting to play an important role in rail transportation system. As a typical public rail transportation system, the Bay Area Rapid Transit (BART) maintains supervision over all phases of its system, including train operations, passenger services, power delivery, and wayside facilities. Big Data analytics is a key element within all of these functions. Schultz in [125] point out that the critical role of BART's operational analytics is ensuring schedule reliability. Using Markov chain model, a multimodal transportation network in London is developed with better information clusters for transportation efficiency [126]. In [127], Big Data analytics is applied in Utrecht, Netherlands to predict the traffic and improve operations with data from mobile phones, smart cards and computers.
铁路运输系统已经通过先进的IT技术进行了改造。它们是大数据分析的主要优势。这是因为铁路运输系统通常是

封闭系统，其执行大量数据的复杂处理，例如实时列车速度和位置，某个站的列车出发和到达时间以及乘客OD信息。大数据分析可以使铁路运输运营商更好地控制列车，提高铁路运输系统的运营效率。在工业领域，大数据分析开始在铁路运输系统中发挥重要作用。作为典型的公共铁路运输系统，湾区快速交通（BART）保持对其系统所有阶段的监督，包括列车运营，客运服务，供电和路旁设施。大数据分析是所有这些功能中的关键要素。Schultz在[125]中指出，BART运营分析的关键作用是确保进度可靠性。利用马尔可夫链模型，开发了伦敦的多式联运网络，为运输效率提供了更好的信息集群[126]。在[127]中，大数据分析应用于荷兰乌得勒支，以预测交通并通过移动电话，智能卡和计算机的数据改善运营。

In academia, substantial work has been done about using Big Data analytics in rail transportation management and control. Using the passenger OD information of Shanghai rail transportation line 1, Jiang et al. [128] evaluate the train timetable efficiency. This method is verified in a real rail transportation system that involves more than 1 million passenger trips and 600 trains. Yin et al. [129] present a smart train operation (STO) method which combines the advantages of automatic train operation (ATO) and manual driving. The fusion of expert knowledge and data mining algorithm is applied in STO method. The results suggest that in energy consumption and riding comfort, this proposed method is better than ATO, and in punctuality and parking accuracy, it is also better than manual driving. Chen et al. [130] propose two simplified models about the relationship between train stop error and the train control parameters by using train speed data, Fig. 5. A typical framework of using Big Data analytics for asset maintenance.
在学术界，在铁路运输管理和控制中使用大数据分析方面已经做了大量工作。利用上海轨道交通1号线的客运OD信息，Jiang等。 [128]评估列车时刻表的效率。这种方法在真实的铁路运输系统中得到验证，该系统涉及超过100万人次的旅客和600列火车。尹等人。 [129]提出了一种智能列车运行（STO）方法，其结合了自动列车运行（ATO）和手动驾驶的优点。将专家知识与数据挖掘算法融合，应用于STO方法。结果表明，在能耗和乘坐舒适性方面，该方法优于ATO，在准时性和停车精度方面，也优于手动驾驶。陈等人。 [130]通过使用列车速度数据，提出了关于列车停止误差与列车控制参数之间关系的两个简化模型，如图5所示。使用大数据分析进行资产维护的典型框架。




time data and distance data before stopping, and introduce one online learning algorithm - polynomial adaline algorithm to increase the parking accuracy. The results show that the proposed simplified models and the online learning algorithms are effective in reducing the parking error and correct the bias of train stop error distribution. Zhou [131] apply two typical machine learning algorithms Gaussian processes and Boosting to improve train stop accuracy by utilizing a number of the initial velocity data and distance data before stopping, the results show that Gaussian process regression algorithm gets the best performance. Hou et al. [132] propose three train stop control algorithms which chooses initial braking position data, braking force data and their combined data as control input. Based on terminal iterative learning control (TILC), these algorithms use the stop position error in previous braking process to improve train stop accuracy. Chen et al. [133] use a new machine learning technique and propose novel online learning control algorithms to realize train automatic stop control. The algorithm includes heuristic online learning algorithm (HOA), gradient-descent based online learning algorithm (GOA), and RL-based online learning algorithm (RLA). The required parameters come from the track-side balises. The results suggest that this method can limit the stopping errors in the range of ±0.30m under regular interferences.
停止前的时间数据和距离数据，并引入一种在线学习算法 - 多项式adaline算法，以提高停车精度。结果表明，所提出的简化模型和在线学习算法可有效地减少停车误差并纠正列车停车误差分布的偏差。周[131]应用两种典型的机器学习算法高斯过程和Boosting，通过利用停止前的一些初始速度数据和距离数据来提高列车停止精度，结果表明高斯过程回归算法获得了最佳性能。侯等人。 [132]提出了三种列车停止控制算法，其选择初始制动位置数据，制动力数据及其组合数据作为控制输入。基于终端迭代学习控制（TILC），这些算法使用先前制动过程中的停止位置误差来提高列车停止精度。陈等人。 [133]使用新的机器学习技术，并提出新颖的在线学习控制算法，以实现列车自动停止控制。该算法包括启发式在线学习算法（HOA），基于梯度下降的在线学习算法（GOA）和基于RL的在线学习算法（RLA）。所需参数来自轨道侧应答器。结果表明，该方法可以在常规干扰下将停止误差限制在±0.30m范围内。

F. Asset Maintenance

F.资产维护

In ITS, there are substantial asset that is dependent on large amounts of data to operate and maintain. Proper asset maintenance approach is very important for protecting ITS capital and reduce maintenance costs. Big Data analytics can help identify problems more quickly and accurately, and minimize maintenance costs. A typical framework of using Big Data analytics for asset maintenance decision making [134] is shown in Fig. 5. Onboard and Infrastructure data is collected from different sensors. Physical failure data such as pavement degradation, ballast aging, track geometry etc. can be used directly. Text data such as experience based information and maintenance report, symbolic description etc., can be processed to extract important information. Infrastructure and vehicle state data such as temperature, humidity, etc. can be processed with data driven method, and obtain the condition indicators. The results from the three process methods is integrated and get the accurate diagnosis of asset condition and determine the remaining useful life of asset, which can be used for end users to make maintenance or operation decisions. One example of Big Data analytics based maintenance is conducted by Dutch railways on Axle Box Acceleration (ABA). With one Terabyte of track degradation data, a self-learning and adapted mechanisms is performed [135]. Thaduri et al. [134] introduce a hybrid modelling approach to provide accurate diagnosis asset condition so as to determine the remaining useful life of asset. The proposed method gives an insight in providing maintenance decision making for end users. Based on semantic data models, a railway asset monitoring system is implemented in [136] and prove to be more capable for data integration, extensibility, and compatibility compared with traditional approaches. Using data collected by multiple inspection vehicles in 330,000 km of railroad track, Zarembski [137] introduce the procedure of data collection, storage and plan the rail track maintenance with Big Data analytics so as to optimize its capital infrastructure and keep costs under control. Using the data from smart phones and GPS co-ordinates, Network Rail in UK successfully improve the track defect position from 1 mile to 5 meters, which significantly reduce the time to fix the rail track [138]. Using huge amount of historical system state data, in combination with train type data, maintenance action data, inspection schedule data, and system failure data, Li et al. [139] explore several machine learning based analytical methods to automatically learn regulations and construct failure estimation models. The models can use the real-time data to estimate if the current conditions will lead to system failure. A bilevel feature extraction-based text mining method is proposed in [140], where features extracted at semantic and syntax levels are used. The proposed method significantly improves the fault diagnosis precision for all fault classes.

在ITS中，有大量资产依赖于大量数据来运营和维护。适当的资产维护方法对于保护ITS资本和降低维护成本非常重要。大数据分析可以帮助更快，更准确地识别问题，并最大限度地降低维护成本。使用大数据分析进行资产维护决策[134]的典型框架如图5所示。从不同的传感器收集板载和基础设施数据。可以直接使用诸如路面退化，压载物老化，轨道几何形状等物理故障数据。可以处理诸如基于经验的信息和维护报告，符号描述等的文本数据以提取重要信息。可以使用数据驱动方法处理诸如温度，湿度等的基础设施和车辆状态数据，并获得状况指标。整合三种过程方法的结果，获得资产状况的准确诊断，确定资产的剩余使用寿命，可供最终用户做出维护或操作决策。基于大数据分析的维护的一个示例由荷兰铁路公司在Axle Box Acceleration（ABA）上进行。利用1TB的轨道退化数据，可以执行自学习和适应机制[135]。Thaduri等人。 [134]引入混合建模方法，以提供准确的诊断资产条件，以确定资产的剩余使用寿命。所提出的方法提供了为最终用户提供维护决策的洞察力。基于语义数据模型，铁路资产监控系统在[136]中实施，并证明与传统方法相比，更能够实现数据集成，可扩展性和兼容性。Zarembski [137]利用多条检查车辆在330,000公里的铁路轨道上收集的数据，介绍了数据采集，存储和使用大数据分析计划铁路轨道维护的程序，以优化其资本基础设施并控制成本。使用来自智能手机和GPS坐标的数据，英国的Network Rail成功地将轨道缺陷位置从1英里提高到5米，这大大缩短了修复铁路轨道的时间[138]。利用大量的历史系统状态数据，结合列车类型数据，维护行动数据，检查计划数据和系统故障数据，Li等人。 [139]探索了几种基于机器学习的分析方法，以自动学习规则并构建故障估计模型。模型可以使用实时数据来估计当前条件是否会导致系统故障。在[140]中提出了一种基于双层特征提取的文本挖掘方法，其中使用了在语义和语法级别提取的特征。该方法显着提高了所有故障类别的故障诊断精度。

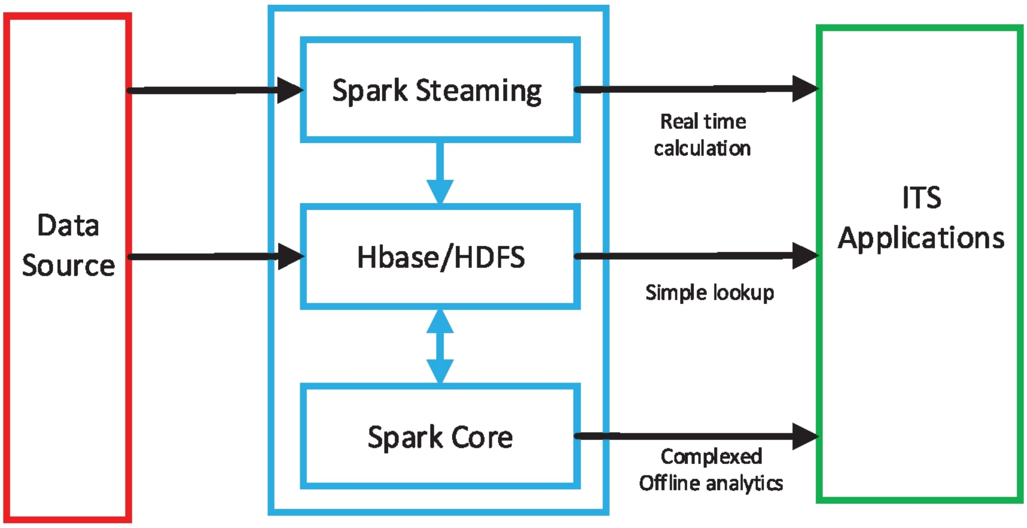
## VI. BIG DATA PLATFORMS IN ITS
## VI。其中的大数据平台

Big Data analytics in ITS have been evolving with the help from advanced Big Data platforms. The Big Data platform leverages distributed file system and parallel computing capability to enable fast data process. It is capable of making sense of Big Data as well as supporting large-scale system optimization.

ITS中的大数据分析在先进的大数据平台的帮助下不断发展。大数据平台利用分布式文件系统和并行计算功能来实现快速数据处理。它能够理解大数据并支持大规模系统优化。

Apache Hadoop is the most popular open source software framework for distributed process and storage of large amount of data sets. Hadoop is a universal Big Data process platform, where various kinds data process or data analytical operations can be carried out. The distributed process capability makes Hadoop well-suited for analyzing the data in ITS, such as smart card data, diverse sensors, social media, GPS data etc. Apache Spark is the latest open-source platform for large amount of data sets processing that peculiarly adapts to machine learning tasks [141]. Spark adopts the same distributed storage technology as Hadoop, and it allows user Fig. 6. A typical framework of using Apache Spark platform in ITS.

Apache Hadoop是最流行的开源软件框架，用于分布式处理和存储大量数据集。Hadoop是一个通用的大数据处理平台，可以执行各种数据处理或数据分析操作。分布式处理能力使Hadoop非常适合分析ITS中的数据，如智能卡数据，各种传感器，社交媒体，GPS数据等。Apache Spark是用于大量数据集处理的最新开源平台，特别适用于机器学习任务[141]。Spark采用与Hadoop相同的分布式存储技术，允许用户使用图6。在ITS中使用Apache Spark平台的典型框架。




programs to load data into a clusters memory and query it repeatedly. Spark is well-suited to machine learning approaches. The Big Data analytics approaches we introduced in the last subsection are machine learning based, and they can definitely be performed in both Hadoop and Spark platforms. The Big Data platform with the data analytics approaches running on it, will play a huge role in Big Data analytics in ITS.

将数据加载到集群内存并重复查询的程序。Spark非常适合机器学习方法。我们在上一小节中介绍的大数据分析方法是基于机器学习的，它们可以在Hadoop和Spark平台上进行定义。运行数据分析方法的大数据平台将在ITS中的大数据分析中发挥重要作用。

A typical framework of using Spark platform in ITS is shown in Fig. 6. Data from different sources are collected by HBase (Hadoop Database) APIs, and they are sent to the data center. Spark Streaming processes the data in real time. Some real time tasks, such as vehicle speed detection, vehicle identification, real time warning etc. can be implemented. HBase is a distributed open source database. It will perform high level feature extraction, and create index for massive data sets, so as to improve the effectiveness and efficiency of data retrieval. Spark Core is the foundation of spark system, and it can carry out off-line tasks with distributed computation capability. Critical tasks such as traffic management and control, accident analysis etc. can be conducted under Spark Core engine.

在ITS中使用Spark平台的典型框架如图6所示。来自不同来源的数据由HBase（Hadoop数据库）API收集，并发送

到数据中心。 Spark Streaming实时处理数据。可以实现一些实时任务，例如车速检测，车辆识别，实时警告等。HBase是一个分布式开源数据库。它将执行高级特征提取，并为海量数据集创建索引，从而提高数据检索的有效性和效率。Spark Core是火花系统的基础，它可以执行具有分布式计算功能的离线任务。可以在Spark Core引擎下进行交通管理和控制，事故分析等关键任务。

Different from the general-purpose Big Data platform, in transportation area, several platforms have been proposed to process the transportation data.
与通用大数据平台不同，在交通领域，已经提出了几个平台来处理运输数据。

Mian et al. [142] propose a platform with multiple engines to support various types of analytic for traffic data. Zareian et al. [143] propose a monitor system named K-Feed for performance analysis of applications deployed on cloud. Shtern et al. [144] propose a conceptual architecture for a data engine, Godzilla, to perform real-time traffic data process and support analytical operation over transportation data. They design a multi-cluster approach to handle large amount of growing data under various kind of workloads and different number of users. Khazaei et al. [145], propose a platform to perform analytical operation on urban transportation data. The platform can be used by traffic-related software developers or directly by traffic engineers and researchers to gain insights of traffic patterns. Chaolong et al. [146] study the development trend of the virtual data center and its technical advantages, proposes a scheme of virtual system of smart transportation data center based on VMware vSphere. A Big Data simulation platform is proposed in [147] for Greater Toronto Area. The platform enables Big Data transportation applications to run in real time.
Mian等。 [142]提出了一个具有多个引擎的平台，以支持各种类型的交通数据分析。Zareian等人。 [143]提出了一个名为K-Feed的监控系统，用于对部署在云上的应用程序进行性能分析。斯特恩等人。 [144]提出了一个数据引擎哥斯拉的概念架构，以执行实时交通数据处理并支持对运输数据的分析操作。他们设计了一种多集群方法，可以处理各种工作负载和不同用户数下的大量增长数据。Khazaei等。 [145]，提出一个平台，对城市交通数据进行分析操作。该平台可供交通相关软件开发人员使用，也可由交通工程师和研究人员直接使用，以获取交通模式的见解。Chaolong等。 [146]研究了虚拟数据中心的发展趋势及其技术优势，提出了一种基于VMware vSphere的智能交通数据中心虚拟系统方案。大多伦多地区[147]提出了一个大数据模拟平台。该平台使大数据运输应用程序能够实时运行。

Real time data streaming process function is a necessary part of Big Data process platform in ITS. Because there are many real-time applications such as traffic monitor and control, and public transportation schedule. Based on the tradition Big Data process system, substantial real-time data streaming systems have been proposed in ITS. Guerreiro et al. [148] propose an ETL (extract, transform and load) architecture for intelligent transportation systems, addressing an application scenario on dynamic toll charging for highways. The proposed architecture is capable of handling real-time and historical data using Big Data technologies such as Spark on Hadoop and MongoDB. A data stream processing platform is proposed in [149], which supports a mechanism for sharing multiparty data sources, software components, and even intermediate results. They give an example of using this platform to conduct traffic management. A comprehensive and flexible architecture based on distributed computing platform for real-time traffic control is proposed in [150]. They have partly realized the architecture in a prototype platform that employs Kafka, a state-of-theart Big Data tool for building data pipelines and stream processing.
实时数据流处理功能是ITS中大数据处理平台的必要组成部分。因为有许多实时应用，如交通监控和公共交通时间表。基于传统的大数据处理系统，ITS中已经提出了大量的实时数据流系统。Guerreiro等。 [148]提出了一种用

于智能交通系统的ETL（提取，转换和加载）架构，解决了高速公路动态收费的应用场景。所提出的架构能够使用大数据技术（如Spark on Hadoop和MongoDB）处理实时和历史数据。在[149]中提出了一种数据流处理平台，它支持共享多方数据源，软件组件甚至中间结果的机制。他们举了一个使用这个平台进行交通管理的例子。在[150]中提出了一种基于分布式计算平台的全面灵活的体系结构，用于实时交通控制。他们部分地在原型平台中实现了架构，该平台采用了Kafka，这是一种用于构建数据管道和流处理的最先进的大数据工具。

Data injection is another critical part of Big Data process system. It is used to transfer data between Big Data process system and relational databases or mainframes. As a popular data injection system, Apache Sqoop has been widely adopted in ITS. For example, Sqoop is used with Hadoop in traffic management system in [151]. It has also been deployed to process vehicle diagnostics data and deliver useful outcomes that can be used by actors in automotive ecosystems [152]. In [153], Apache Sqoop is used to ingest ITS relational data. Apache Flume is another popular data injection system that processes unstructured data, and it has been adopted to process log data in ITS [154].
数据注入是大数据处理系统的另一个关键部分。它用于在大数据处理系统和关系数据库或大型机之间传输数据。作为一种流行的数据注入系统，Apache Sqoop已在ITS中得到广泛采用。例如，在[151]中，Sqoop与交通管理系统中的Hadoop一起使用。它还被用于处理车辆诊断数据并提供可供汽车生态系统中的参与者使用的有用结果[152]。在[153]中，Apache Sqoop用于摄取ITS关系数据。Apache Flume是另一种处理非结构化数据的流行数据注入系统，它已被用于处理ITS中的日志数据[154]。

# VII. OPEN CHALLENGES
# 七。开放的挑战

Although Big Data analytics has made great achievements in ITS, there are still substantial open challenges have not been fully studied. They need to be tackled in future works. This section introduce the main open challenges of using Big Data analytics in ITS as follows.
尽管大数据分析在ITS方面取得了巨大成就，但仍有大量的开放性挑战尚未得到充分研究。他们需要在未来的工作中得到解决。本节介绍在ITS中使用大数据分析的主要开放挑战如下。

• Data collection: Due to the frequent movement of vehicles and pedestrians, data collected in transportation may be inaccurate, incomplete or unreliable in particular locations or at certain times. For instance, not all vehicles are embedded with the techniques needed to provide real-time location data, and road traffic data from road sensors can be missing. One possible way to tackle the challenger is to invest new data collection technologies and improve the data collection capability. With the development IoT, new sensor techniques are invented annually, which can help improve data collection and data quality. In addition, the adoption of data capturing automation to minimize manual data entry is also essential to data quality improvement.
•数据收集：由于车辆和行人频繁移动，在特定地点或特定时间运输中收集的数据可能不准确，不完整或不可靠。例如，并非所有车辆都嵌入了提供实时位置数据所需的技术，并且可能缺少来自道路传感器的道路交通数据。应对挑战者的一种可能方法是投资新的数据收集技术并提高数据收集能力。随着物联网的发展，每年都会发明新的传感器技术，这有助于提高数据收集和数据质量。此外，采用数据捕获自动化以最小化手动数据输入对于数据质量改进也是必不可少的。

• Data privacy: In the era of Big Data, the most challenging and concerned problem is privacy [155]. Personal privacy may be leaked during data transmission, storage and usage [156]. Data collected from transportation systems

•数据隐私：在大数据时代，最具挑战性和最关注的问题是隐私[155]。在数据传输，存储和使用过程中可能会泄露个人隐私[156]。从运输系统收集的数据

used to be non-personal data, such as vehicle location, traffic flow data. However, privacy problems have been concerned since personal data collection by the public and private sectors grows over time. For example, the location of individuals and vehicles can be easily collected. If these data are not strictly protected, people who steal these data would harm the owner of the data. Therefore, privacy protection is an important thing for Big Data applications in ITS. To prevent unauthorized disclosure of the personal private information, governments should develop complete data privacy laws which include what data can be published, the scope of the data publishing and using, the basic principles of data distribution, data availability and other areas [157]. The transportation departments should strictly regulate the personal data definition, strengthen the management of data security certification, and use more advanced algorithms to improve the data security level.
曾经是非个人数据，例如车辆位置，交通流量数据。然而，由于公共和私营部门的个人数据收集随着时间的推移而增长，因此隐私问题一直受到关注例如，可以容易地收集个人和车辆的位置。如果这些数据没有受到严格保护，窃取这些数据的人将会损害数据所有者。因此，隐私保护对于ITS中的大数据应用来说是一件重要的事情。为防止未经授权披露个人隐私信息，政府应制定完整的数据隐私法，包括可以发布的数据，数据发布和使用的范围，数据分发的基本原则，数据可用性和其他方面[157]。运输部门应严格规范个人数据定义，加强数据安全认证管理，采用更先进的算法提高数据安全水平。

• Data storage: Currently, the data volume has jumped from TB level to PB level, and the growth in data storage capacity is far behind the data growth. Especially in ITS, it will produce a variety of data from the various sensors every day. Traditional data storage infrastructure and database tools have been unable to cope with the increasingly large and complex mass data [158]. Therefore, designing the most reasonable data storage architecture has become a key challenge. The main public cloud storage providers, such as Google and Microsoft, continue to improve their services with integrated Big Data capabilities, and multi-cloud storage and hybrid storage are emerging as key areas for Big Data storage. Their compute bursting capabilities have advantages in many forms of compute-intensive analytics workloads. In addition, combining intelligence with storage is also a good solution. Enterprises are looking for smart management tools which can provide integrated analytics within storage. This enables them to conduct resource monitoring and make full use of storage infrastructure. • Data processing: Timeliness is crucial to Big Data applications in ITS, these applications include traffic data preprocessing, traffic state recognition, real-time traffic control, dynamic route guidance and real-time bus scheduling. Traffic data which contain different formats from diverse sources, must be compared with the historical data, then processed within a short time [159]. The data processing system must be able to process more complicated and increasingly expanding data. How to guarantee the process timeliness with so large and fast data is a big challenge. Many general Big Data frameworks that handle real time data sources, such as Apache Storm, Apache Flink, Apache Samza, Apache Spark Streaming and Kafka Streams, have appeared recently. In addition, dedicated Big Data processing frameworks for ITS have also been developed, such as platform for real-time traffic control, and estimating the average speed and the congested sections of a highway. These processing framework provide good solutions to real time data processing.
•数据存储：目前，数据量已从TB级别跃升至PB级别，数据存储容量的增长远远落后于数据增长。特别是在ITS中，它每天都会从各种传感器产生各种数据。传统的数据存储基础设施和数据库工具无法应对日益庞大和复杂的海量数据[158]。因此，设计最合理的数据存储架构已成为一项关键挑战。主要的公共云存储提供商（如Google和

Microsoft）通过集成的大数据功能继续改进其服务，多云存储和混合存储正在成为大数据存储的关键领域。他们的计算突发功能在许多形式的计算密集型分析工作负载中具有优势。此外，将智能与存储相结合也是一个很好的解决方案。企业正在寻找能够在存储中提供集成分析的智能管理工具。这使他们能够进行资源监控并充分利用存储基础架构。 •数据处理：及时性对ITS中的大数据应用至关重要，这些应用包括交通数据预处理，交通状态识别，实时交通控制，动态路线引导和实时总线调度。包含来自不同来源的不同格式的交通数据必须与历史数据进行比较，然后在短时间内处理[159]。数据处理系统必须能够处理更复杂和不断扩展的数据。如何通过如此大而快的数据保证过程的及时性是一个巨大的挑战。最近出现了许多处理实时数据源的通用大数据框架，例如Apache Storm，Apache Flink，Apache Samza，Apache Spark Streaming和Kafka Streams。此外，还开发了专用于ITS的大数据处理框架，例如用于实时交通控制的平台，以及估算高速公路的平均速度和拥挤路段。这些处理框架为实时数据处理提供了良好的解决方案。

• Data opening: To enable transportation service users and App developers to find and re-use data effectively, data need to be archived and made publicly accessible in good quality. Data quality refers to its accuracy, completeness, reliability, and consistency [160], [161]. Without good data quality, Big Data will be misleading to decisionmaking and even produce harmful results. However, opening up data with good quality might require time and money. There is a trade-off between opening up data quickly at low cost and making high quality data available at high costs, which makes opening up good quality data one more big challenge. Effective solutions include the adoption of automatic data capturing and/or utilization of artificial intelligence to verify the data. Additionally, the transportation departments should have a data management process enacted to ensure pristine and accurate data.

•数据开放：为了使运输服务用户和应用程序开发人员能够有效地发现和重新使用数据，需要将数据存档并以高质量公开访问。数据质量是指其准确性，完整性，可靠性和一致性[160]，[161]。如果没有良好的数据质量，大数据将会误导决策，甚至产生有害的结果。但是，打开高质量的数据可能需要时间和金钱。在以低成本快速打开数据和以高成本提供高质量数据之间存在权衡，这使得打开高质量数据成为一个更大的挑战。有效的解决方案包括采用自动数据捕获和/或利用人工智能来验证数据。此外，运输部门应制定数据管理流程，以确保原始和准确的数据。

# VIII. CONCLUSIONS

# 八。结论

In this paper, we presented the development of Big Data and the relevant knowledge of ITS. The framework of conducting Big Data analytics in ITS was discussed. We summarized the data source and collection methods, data analytics methods and platforms, and Big Data analytics application categories in ITS. We presented several applications of Big Data analytics in ITS, including asset maintenance, road traffic flow prediction, road traffic accidents analysis, public transportation service planning, personal travel route planning and rail transportation management and control. Several open challenges of using Big Data analytics in ITS were discussed in this paper, including data collection, data privacy, data storage, data processing, and data opening. Big Data analytics will have profound impacts on the design of intelligent transportation system, and make it safer, more efficient and profitable.

在本文中，我们介绍了大数据的发展和ITS的相关知识。讨论了在ITS中进行大数据分析的框架。我们总结了ITS中的数据源和收集方法，数据分析方法和平台以及大数据分析应用程序类别。我们在ITS中介绍了大数据分析的几种应用，包括资产维护，道路交通流量预测，道路交通事故分析，公共交通服务规划，个人旅行路线规划以及铁路运输管理和控制。本文讨论了在ITS中使用大数据分析的几个公开挑战，包括数据收集，数据隐私，数据存

储，数据处理和数据打开。大数据分析将对智能交通系统的设计产生深远影响，使其更安全，更高效，更有效。

# REFERENCES
# 参考

[1] G. Bello-Orgaz, J. J. Jung, and D. Camacho, “Social big data: Recent achievements and new challenges,” Inf. Fusion, vol. 28, pp. 45–59, Mar. 2016.
[1] G. Bello-Orgaz，J。J. Jung和D. Camacho，“社会大数据：近期成就和新挑战”，Inf。融合，第一卷28，pp.45-59，2016年3月。

[2] M. Chen, S. Mao, and Y. Liu, “Big data: A survey,” Mobile Netw. Appl., vol. 19, no. 2, pp. 171–209, Apr. 2014.
[2] M. Chen，S。Mao和Y. Liu，“大数据：一项调查，”移动网络。 Appl。，vol。 19，没有。 2，pp.171-209，2014年4月。

[3] H. Chen, R. H. Chiang, and V. C. Storey, “Business intelligence and analytics: From big data to big impact,” MIS Quart., vol. 36, no. 4, pp. 1165–1188, 2012.
[3] H. Chen，R。H. Chiang和V. C. Storey，“商业智能与分析：从大数据到重大影响”，MIS Quart。，vol。 36，不。 4，pp.1165-1188,2012。

[4] T. B. Murdoch and A. S. Detsky, “The inevitable application of big data to health care,” JAMA, vol. 309, no. 13, pp. 1351–1352, 2013.
[4] T. B. Murdoch和A. S. Detsky，“大数据不可避免地应用于医疗保健”，JAMA，第一卷。 309，没有。 13，pp.1351-1352,2013。

[5] M. Mayilvaganan and M. Sabitha, “A cloud-based architecture for bigdata analytics in smart grid: A proposal,” in Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC), Dec. 2013, pp. 1–4.
[5] M. Mayilvaganan和M. Sabitha，“基于云的智能电网中的大数据分析架构：提案”，Proc。 IEEE Int。 CONF。 COMPUT。 INTELL。 COMPUT。 RES。 （ICCIC），2013年12月，第1-4页。

[6] L. Qi, “Research on intelligent transportation system technologies and applications,” in Proc. Workshop Power Electron. Intell. Transp. Syst., 2008, pp. 529–531.
[6] L. Qi，“智能交通系统技术与应用研究”，载于Proc。车间电力电子。 INTELL。运输。 Syst。，2008，pp.529-531。

[7] S.-H. An, B.-H. Lee, and D.-R. Shin, “A survey of intelligent transportation systems,” in Proc. Int. Conf. Comput. Intell., Jul. 2011, pp. 332–337.
[7] S.-H. An，B.-H。李和D.-R. Shin，“智能交通系统调查”，Proc。诠释。 CONF。 COMPUT。 Intell。，2011年7月，第332-337页。

[8] N.-E. El Faouzi, H. Leung, and A. Kurian, “Data fusion in intelligent transportation systems: Progress and challenges—A survey,” Inf. Fusion, vol. 12, no. 1, pp. 4–10, 2011.
[8] N.-E. El Faouzi，H。Leung和A. Kurian，“智能交通系统中的数据融合：进展与挑战 - 一项调查”，Inf。融合，第一卷12，不。 1，pp.4-10,2011。

[9] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, “Datadriven intelligent transportation systems: A

survey," IEEE Trans. Intell. Transp. Syst., vol. 12, no. 4, pp. 1624–1639, Dec. 2011.
[9] J. Zhang，F.-Y。 Wang，K。Wang，W.-H。 Lin，X。Xu和C. Chen，“Datadriven智能交通系统：一项调查”，IEEE Trans。 INTELL。运输。 Syst。，vol。 12，不。 4，pp.1644-1639，2011年12月。

[10] Q. Shi and M. Abdel-Aty, “Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways,” Transp. Res. C, Emerg. Technol., vol. 58, pp. 380–394, Sep. 2015.
[10] Q. Shi和M. Abdel-Aty，“实时交通运营中的大数据应用以及城市高速公路的安全监测和改进”，运输。 RES。 C，Emerg。 Technol。，vol。 58，pp.380-394，2015年9月。

[11] N. Mohamed and J. Al-Jaroodi, “Real-time big data analytics: Applications and challenges,” in Proc. Int. Conf. High Perform. Comput. Simulation, Jul. 2014, pp. 305–310.
[11] N. Mohamed和J. Al-Jaroodi，“实时大数据分析：应用与挑战”，Proc。诠释。 CONF。高绩效。 COMPUT。 Simulation，2014年7月，第305-310页。

[12] X. Lin, P. Wang, and B. Wu, “Log analysis in cloud computing environment with hadoop and spark,” in Proc. 5th IEEE Int. Conf. Broadband Netw. Multimedia Technol. (IC-BNMT), Nov. 2013, pp. 273–276.
[12] X. Lin，P。Wang和B. Wu，“使用hadoop和spark进行云计算环境中的日志分析”，Proc。第五届IEEE国际CONF。宽带网。多媒体技术（IC-BNMT），2013年11月，第273-276页。

[13] M. Zaharia et al., “Fast and interactive analytics over Hadoop data with spark,” USENIX Login, vol. 37, no. 4, pp. 45–51, 2012.
[13] M. Zaharia等人，“使用spark对Hadoop数据进行快速和交互式分析”，USENIX登录，第一卷。 37，不。 4，pp.45-51,2012。

[14] D. Corrigan, P. Zikopoulos, K. Parasuraman, T. Deutsch, D. Deroos, and J. Giles, Harness the Power of Big Data the IBM Big Data Platform. 1st ed. New York, NY, USA: McGraw-Hill, Nov. 2012.
[14] D. Corrigan，P。Zikopoulos，K。Parasuraman，T。Deutsch，D。Deroos和J. Giles，Harness the Power of Big Data the IBM Big Data Platform。第1版。纽约，纽约，美国：McGraw-Hill，2012年11月。

[15] L. Basche, “Says solving ‘big data’ challenge involves more than just managing volumes of data,” Bus. Wire, San Francisco, CA, USA, Tech. Rep., Jun. 2011.
[15] L. Basche，“解决'大数据'挑战的问题不仅仅涉及管理大量数据，”巴士。电线，旧金山，加利福尼亚州，美国，科技。众议员，2011年6月。

[16] M. Bagchi and P. R. White, “The potential of public transport smart card data,” Transp. Policy, vol. 12, no. 5, pp. 464–474, 2005.
[16] M. Bagchi和P. R. White，“公共交通智能卡数据的潜力”，运输。政策，第一卷12，不。 5，pp.464-474,2005。

[17] M.-P. Pelletier, M. Trépanier, and C. Morency, “Smart card data use in public transit: A literature review,” Transp. Res. C, Emerg. Technol., vol. 19, no. 4, pp. 557–568, 2011.
[17] M.-P. Pelletier，M.Trépanier和C. Morency，“公共交通中的智能卡数据使用：文献综述”，运输。 RES。 C，Emerg。 Technol。，vol。 19，没有。 4，pp.557-568,2011。

[18] Y. Liu, X. Weng, J. Wan, X. Yue, and H. Song, “Exploring data validity in transportation systems for smart cities,”

IEEE Commun. Mag., vol. 55, no. 5, pp. 26–33, 2017.
[18] Y. Liu，X。Weng，J。Wan，X。Yue和H. Song，“探索智能城市交通系统的数据有效性”，IEEE Commun。Mag。，vol。 55，不。 5，pp.26-33,2017。

[19] H. Nishiuchi, J. King, and T. Todoroki, “Spatial-temporal daily frequent trip pattern of public transport passengers using smart card data,” Int. J. Intell. Transp. Syst. Res., vol. 11, no. 1, pp. 1–10, 2013.
[19] H. Nishiuchi，J。King和T. Todoroki，“使用智能卡数据的公共交通乘客的时空日常频繁旅行模式”，Int。 J. Intell。运输。 SYST。 Res。，vol。 11，不。 1，pp.1-10,2013。

[20] M. A. Munizaga and C. Palma, “Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile,” Transp. Res. C, Emerg. Technol., vol. 24, pp. 9–18, Oct. 2012.
[20] M. A. Munizaga和C. Palma，“从智利圣地亚哥的无源智能卡数据估算分解的多式联运公共交通源 - 目的地矩阵”，运输。 RES。 C，Emerg。 Technol。，vol。 24，pp.9-18，2012年10月。

[21] K. A. Chu and R. Chapleau, “Enriching archived smart card transaction data for transit demand modeling,” Transp. Res. Rec., J. Transp. Res. Board, pp. 63–72, Dec. 2008.
[21] K. A. Chu和R. Chapleau，“丰富存档需求建模的存档智能卡交易数据”，运输。 RES。 Rec。，J。Transp。 RES。董事会，第63-72页，2008年12月。

[22] M. A. Munizaga and C. Palma, “Estimation of a disaggregate multimodal public transport origin–destination matrix from passive smartcard data from santiago, chile,” Transp. Res. C, Emerg. Technol., vol. 24, pp. 9–18, 2012.
[22] M. A. Munizaga和C. Palma，“从圣地亚哥，智利的被动智能卡数据估算分解的多式联运公共交通起始目的地矩阵”，运输。 RES。 C，Emerg。 Technol。，vol。 24，pp.9-18,2012。

[23] H. Gong, C. Chen, E. Bialostozky, and C. T. Lawson, “A GPS/GIS method for travel mode detection in New York City,” Comput., Environ. Urban Syst., vol. 36, no. 2, pp. 131–139, 2012.
[23] H. Gong，C。Chen，E。Bialostozky和C. T. Lawson，“用于纽约市旅行模式检测的GPS / GIS方法”，Comput。，Environ。 Urban Syst。，vol。 36，不。 2，pp.131-139,2012。

[24] X. Wang, S. Zhao, and L. Dong, “Research and application of traffic visualization based on vehicle GPS big data,” in Proc. Int. Conf. Intell. Transp., 2016, pp. 293–302.
[24] X. Wang，S。Zhao和L. Dong，“基于车辆GPS大数据的交通可视化的研究和应用”，Proc。诠释。 CONF。 INTELL。 Transp。，2016，pp.293-302。

[25] C. Asensio, J. López, R. Pagán, I. Pavón, and M. Ausejo, “GPSbased speed collection method for road traffic noise mapping,” Transp. Res. D, Transp. Environ., vol. 14, no. 5, pp. 360–366, 2009.
[25] C. Asensio，J.López，R.Pagán，I.Pavón和M. Ausejo，“基于GPS的道路交通噪声测绘速度收集方法”，运输。 RES。 D，运输。环境，卷。 14，没有。 5，pp.360-366,2009。

[26] J. C. Herrera, D. B. Work, R. Herring, X. Ban, Q. Jacobsond, and A. M. Bayen, “Evaluation of traffic data obtained via GPS-enabled mobile phones: The mobile century field experiment,” Transp. Res. C, Emerg. Technol., vol. 18, no. 4, pp. 568–583, Aug. 2010.
[26] J. C. Herrera，D。B. Work，R。Herring，X。Ban，Q。Jacobsond和A. M. Bayen，“通过具有GPS功能的移动电话获得的交通数据评估：移动世纪现场实验”，运输。 RES。 C，Emerg。 Technol。，vol。 18，不。 4，

pp.568-583，2010年8月。

[27] K. G. Courage, M. Doctor, S. Maddula, and R. Surapaneni, “Video image detection for traffic surveillance and control,” Transp. Res. Center, Univ. Florida, Gainesville, FL, USA, Tech. Rep. TD100:FL96119, Mar. 1996.
[27] K. G. Courage，M。Doctor，S。Maddula和R. Surapaneni，“用于交通监视和控制的视频图像检测”，运输。RES。中心，大学佛罗里达州，盖恩斯维尔，佛罗里达州，美国，科技。 Rep.TD100：FL96119，1996年3月。

[28] C. Grant, B. Gillis, and R. Guensler, “Collection of vehicle activity data by video detection for use in transportation planning,” J. Intell. Transp. Syst., vol. 5, no. 4, pp. 343–361, 2000.
[28] C. Grant，B。Gillis和R. Guensler，“通过视频检测收集车辆活动数据，用于交通规划”，J。Intell。运输。Syst。，vol。 5，不。 4，pp.343-361,2000。

[29] M. Kadaieaswaran, V. Arunprasath, and M. Karthika, “Big data solution for improving traffic management system with video processing,” Int. J. Eng. Sci., vol. 7, no. 2, 2017.
[29] M. Kadaieaswaran，V。Arunprasath和M. Karthika，“通过视频处理改善交通管理系统的大数据解决方案”，Int。 J. Eng。 Sci。，vol。 7，不。 2017年2月2日。

[30] J. Lopes, J. Bento, E. Huang, C. Antoniou, and M. Ben-Akiva, “Traffic and mobility data collection for real-time applications,” in Proc. IEEE Intell. Transp. Syst. (ITSC), Sep. 2010, pp. 216–223.
[30] J. Lopes，J。Bento，E。Huang，C。Antoniou和M. Ben-Akiva，“Traf fi c and mobility data collection for real-time applications”，Proc。 IEEE Intell。运输。 SYST。 （ITSC），2010年9月，第216-223页。

[31] C. Antoniou, R. Balakrishna, and H. Koutsopoulos, “Emerging data collection technologies and their impact on traffic management applications,” in Proc. 10th Int. Conf. Appl. Adv. Technol. Transp., Athens, Greece, 2008.
[31] C. Antoniou，R。Balakrishna和H. Koutsopoulos，“新兴数据收集技术及其对交通管理应用的影响”，Proc。第十届国际CONF。申请进阶TECHNOL。运输，雅典，希腊，2008年。

[32] E. Huang, “Algorithmic and implementation aspects of on-line calibration of dynamic traffic assignment,” Ph.D. dissertation, Dept. Civil, Environ. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 2010.
[32] E. Huang，“动态交通分配在线校准的算法和实现方面”，博士。学位论文，土木与环境部。 Eng。，Massachusetts Inst。 Technol。，Cambridge，MA，USA，2010。

[33] E. Uhlemann, “Autonomous vehicles are connecting… [connected vehicles],” IEEE Veh. Technol. Mag., vol. 10, no. 2, pp. 22–25, Jun. 2015.
[33] E. Uhlemann，“自动驾驶车辆正在连接...... [联网车辆]，”IEEE Veh。 TECHNOL。 Mag。，vol。 10，不。 2，pp.22-25，Jun.2015。

[34] C. Chen, T. H. Luan, X. Guan, N. Lu, and Y. Liu. (2017). “Connected vehicular transportation: Data analytics and traffic-dependent networking.” [Online]. Available: https://arxiv.org/abs/1704.08125
[34] C. Chen，T。H. Luan，X。Guan，N。Lu和Y. Liu。 （2017年）。 “连接车辆运输：数据分析和交通依赖网络。”[在线]。可用：https：//arxiv.org/abs/1704.08125

[35] X. Wang, C. Wang, J. Zhang, M. Zhou, and C. Jiang, “Improved rule installation for real-time query service in software-defined Internet of vehicles,” IEEE Trans. Intell. Transp. Syst., vol. 18, no. 2, pp. 225–235, Feb. 2017.
[35] X. Wang，C。Wang，J。Zhang，M。Zhou和C. Jiang，“用于软件定义的车辆互联网中的实时查询服务的改

进规则安装”，IEEE Trans。 INTELL。运输。 Syst。 ，vol。 18，不。 2，pp.225-235，2017年2月。

[36] J. Hu, L. Kong, W. Shu, and M.-Y. Wu, “Scheduling of connected autonomous vehicles on highway lanes,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), Dec. 2012, pp. 5556–5561.
[36] J. Hu，L。Kong，W。Shu和M.-Y.吴，“公路车道上连通自动驾驶汽车的调度”，Proc。 IEEE全球通讯。 CONF。 （GLOBECOM），2012年12月，第5556-5561页。

[37] R. King, “Traffic management in a connected or autonomous vehicle environment,” in Proc. Auto. Passenger Veh., May 2015, pp. 1–20.
[37] R. King，“交通或自主车辆环境中的交通管理”，Proc。汽车。 Passenger Veh。 ，2015年5月，第1-20页。

[38] P. Bedi and V. Jindal, “Use of big data technology in vehicular ad-hoc networks,” in Proc. Int. Conf. Adv. Comput., Commun. Inform. (ICACCI), Sep. 2014, pp. 1677–1683.
[38] P. Bedi和V. Jindal，“在车载ad-hoc网络中使用大数据技术”，Proc。诠释。 CONF。进阶计算，共同。通知。 （ICACCI），2014年9月，第1677-1683页。

[39] J. Contreras-Castillo, S. Zeadally, and J. A. G. Ibañez, “Solving vehicular ad hoc network challenges with big data solutions,” IET Netw., vol. 5, no. 4, pp. 81–84, Jul. 2016.
[39] J. Contreras-Castillo，S。Zeadally和J. A.G.Ibañez，“用大数据解决方案解决车载ad hoc网络挑战”，IET Netw。 ，vol。 5，不。 4，pp.81-84，2016年7月。

[40] M. Tan, I. W. Tsang, and L. Wang, “Towards ultrahigh dimensional feature selection for big data,” J. Mach. Learn. Res., vol. 15, pp. 1371–1429, Apr. 2014.
[40] M. Tan，I。W. Tsang和L. Wang，“迈向大数据的超高维特征选择”，J。Mach。学习。 Res。 ，vol。 15，pp.1371-1429，2014年4月。

[41] A. Mahajan and A. Kaur, “Predictive urban traffic flow model using vehicular big data,” Indian J. Sci. Technol., vol. 9, no. 42, pp. 1–8, 2016.
[41] A. Mahajan和A. Kaur，“使用车辆大数据的预测性城市交通流模型”，印度J. Sci。 Technol。 ，vol。 9，不。 42，pp.1-8,2016。

[42] H. A. Najada and I. Mahgoub, “Anticipation and alert system of congestion and accidents in vanet using big data analysis for intelligent transportation systems,” in Proc. IEEE Symp. Ser. Comput. Intell. (SSCI), Dec. 2016, pp. 1–8.
[42] H. A. Najada和I. Mahgoub，“在智能交通系统中使用大数据分析的虚拟网络中的拥堵和事故的预期和警报系统”，Proc。 IEEE Symp。序列。 COMPUT。 INTELL。 （SSCI），2016年12月，第1-8页。

[43] L. Gong, X. Liu, L. Wu, and Y. Liu, “Inferring trip purposes and uncovering travel patterns from taxi trajectory data,” Cartography Geograph. Inf. Sci., vol. 43, no. 2, pp. 103–114, 2016.
[43] L. Gong，X。Liu，L。Wu和Y. Liu，“从出租车轨迹数据中推断旅行目的并揭示旅行模式”，制图地理。天道酬勤。 Sci。 ，vol。 43，不。 2，pp.103-114,2016。

[44] C. Kang, Y. Liu, X. Ma, and L. Wu, “Towards estimating urban population distributions from mobile call data,” J. Urban Technol., vol. 19, no. 4, pp. 3–21, 2012.
[44] C. Kang，Y。Liu，X。Ma和L. Wu，“从移动通话数据估算城市人口分布”，J。Urban Technol。 ，vol。 19，没有。 4，pp.3-21,2012。

[45] C. Chen, J. Ma, Y. Susilo, Y. Liu, and M. Wang, “The promises of big data and small data for travel behavior (aka human mobility) analysis,” Transp. Res. C, Emerg. Technol., vol. 68, pp. 285–299, Jul. 2016.
[45] C. Chen，J。Ma，Y。Susilo，Y。Liu和M. Wang，“旅行行为（又称人类流动性）分析的大数据和小数据的承诺”，运输。 RES。 C，Emerg。 Technol。 ，vol。 68，pp.285-299，Jul。2016。

[46] J. Zeyu, Y. Shuiping, Z. Mingduan, C. Yongqiang, and L. Yi, “Model study for intelligent transportation system with big data,” Proc. Comput. Sci., vol. 107, pp. 418–426, 2017.
[46] J. Zeyu，Y。Shuiping，Z。Mingduan，C。Yongqiang和L. Yi，“大数据智能交通系统的模型研究”，Proc。 COMPUT。 Sci。 ，vol。 107，pp.418-426,2017。

[47] S. Liu, H. Cao, L. Li, and M. C. Zhou, “Predicting stay time of mobile users with contextual information,” IEEE Trans. Automat. Sci. Eng., vol. 10, no. 4, pp. 1026–1036, Oct. 2013.
[47] S. Liu，H。Cao，L。Li和M. C. Zhou，“用上下文信息预测移动用户的停留时间”，IEEE Trans。自动售货机。科学。 Eng。 ，vol。 10，不。 4，pp.1026-1036，2013年10月。

[48] A. Gal-Tzur, S. M. Grant-Muller, T. Kuflik, E. Minkov, S. Nocera, and I. Shoor, “The potential of social media in delivering transport policy goals,” Transp. Policy, vol. 32, pp. 115–123, Mar. 2014.
[48] A. Gal-Tzur，S。M. Grant-Muller，T。Ku fl ik，E。Minkov，S。Nocera和I. Shoor，“社交媒体在实现交通政策目标方面的潜力”，运输。政策，第一卷32，pp.115-123，2014年3月。

[49] F. Alesiani, K. Gkiotsalitis, and R. Baldessari, “A probabilistic activity model for predicting the mobility patterns of homogeneous social groups based on social network data,” in Proc. 93rd Annu. Meeting Transp. Res. Board, 2014.
[49] F. Alesiani，K。Gkiotsalitis和R. Baldessari，“基于社会网络数据预测同质社会群体流动模式的概率活动模型”，Proc。 93年Annu。会议运输。 RES。董事会，2014年。

[50] Y. Chen, A. Frei, and H. Mahmassani, “From personal attitudes to public opinion: Information diffusion in social networks toward sustainable transportation,” Transp. Res. Rec., J. Transp. Res. Board, vol. 2, pp. 28–37, Nov. 2014.
[50] Y. Chen，A。Frei和H. Mahmassani，“从个人对公众舆论的态度：社会网络中向可持续交通的信息传播”，运输。 RES。 Rec。 ，J。Transp。 RES。董事会，第一卷2，pp.28-37，2014年11月。

[51] B. Pender, G. Currie, A. Delbosc, and N. Shiwakoti, “Social media use during unplanned transit network disruptions: A review of literature,” Transp. Rev., vol. 34, no. 4, pp. 501–521, 2014.
[51] B. Pender，G。Currie，A。Delbosc和N. Shiwakoti，“在计划外交通网络中断期间使用社交媒体：文献综述”，运输。 Rev.，vol。 34，不。 4，pp.501-521,2014。

[52] X. Zheng et al., “Big data for social transportation,” IEEE Trans. Intell. Transp. Syst., vol. 17, no. 3, pp. 620–630, Mar. 2016.
[52] X. Zheng等人，“社会交通的大数据”，IEEE Trans。 INTELL。运输。 Syst。 ，vol。 17，不。 3，pp.620-630，2016年3月。

[53] C. D. Cottrill and S. Derrible, “Leveraging big data for the development of transport sustainability indicators,” J. Urban Technol., vol. 22, no. 1, pp. 45–64, 2015.
[53] C. D. Cottrill和S. Derrible，“利用大数据制定运输可持续性指标”，J。Urban Technol。 ，vol。 22，不。 1，pp.45-64,2015。

[54] G. R. Grob, “Future transportation with smart grids & sustainable energy,” in Proc. 6th Int. Multi-Conf. Syst., Signals Devices (SSD), 2009, pp. 1–5.
[54] G. R. Grob，“智能电网与可持续能源的未来交通”，Proc。第六届国际多屏的配置Syst。，Signals Devices（SSD），2009，pp.1-5。

[55] L. Zhu, F. R. Yu, B. Ning, and T. Tang, “Cross-layer handoff design in MIMO-enabled WLANs for communication-based train control (CBTC) systems,” IEEE J. Sel. Areas Commun., vol. 30, no. 4, pp. 719–728, May 2012.
[55] L. Zhu，F。R. Yu，B。Ning和T. Tang，“基于MIMO的WLAN中基于通信的列车控制（CBTC）系统的跨层切换设计”，IEEE J. Sel。地区Commun。，vol。 30，不。 4，pp.719-728，2012年5月。

[56] L. Zhu, F. R. Yu, B. Ning, and T. Tang, “Handoff performance improvements in MIMO-enabled communication-based train control systems,” IEEE Trans. Intell. Transp. Syst., vol. 13, no. 2, pp. 582–593, Jun. 2012.
[56] L. Zhu，F。R. Yu，B。Ning和T. Tang，“基于MIMO的通信型列车控制系统的切换性能改进”，IEEE Trans。INTELL。运输。 Syst。，vol。 13，不。 2，pp.582-593，2012年6月。

[57] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell, Machine Learning: An Artificial Intelligence Approach. New York, NY, USA: Springer, 2013.
[57] R. S. Michalski，J。G. Carbonell和T. M. Mitchell，机器学习：一种人工智能方法。纽约，纽约，美国：Springer，2013年。

[58] G. A. Seber and A. J. Lee, Linear Regression Analysis, vol. 936. Hoboken, NJ, USA: Wiley, 2012.
[58] G. A. Seber和A. J. Lee，线性回归分析，第一卷。 936.美国新泽西州霍博肯：Wiley，2012。

[59] H. Sun, H. Liu, H. Xiao, R. He, and B. Ran, “Use of local linear regression model for short-term traffic forecasting,” Transp. Res. Rec., J. Transp. Res. Board, pp. 143–150, Jan. 2003.
[59] H. Sun，H。Liu，H。Xiao，R。He和B. Ran，“使用局部线性回归模型进行短期交通预测”，运输。 RES。Rec。，J。Transp。 RES。董事会，第143-150页，2003年1月。

[60] Z. Shan, D. Zhao, and Y. Xia, “Urban road traffic speed estimation for missing probe vehicle data based on multiple linear regression model,” in Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC), Oct. 2013, pp. 118–123.
[60] Z. Shan，D。Zhao和Y. Xia，“基于多元线性回归模型的缺失探测车辆数据的城市道路交通速度估计”，Proc。第16届国际IEEE会议INTELL。运输。 SYST。 （ITSC），2013年10月，第118-123页。

[61] N. Zenina and A. Borisov, “Regression analysis for transport trip generation evaluation,” Inf. Technol. Manage. Sci., vol. 16, no. 1, pp. 89–94, 2013.
[61] N. Zenina和A. Borisov，“运输旅行产生评估的回归分析”，Inf。 TECHNOL。管理。 Sci。，vol。 16，不。 1，pp.89-94,2013。

[62] J. R. Quinlan, “Induction of decision trees,” Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986.
[62] J. R. Quinlan，“决策树的归纳”，马赫。学习。，第一卷。 1，不。 1，pp.81-106,1986。

[63] H. J. Payne and S. Tignor, “Freeway incident detection algorithms based on decision trees with states,” Transp. Res. Rec., vol. 682, pp. 30–37, Jan. 1978.
[63] H. J. Payne和S. Tignor，“基于具有状态的决策树的高速公路事件检测算法”，运输。 RES。 Rec。，vol。

682，pp.30-37，1978年1月。

[64] J. AbelláN, G. LóPez, and J. De OñA, “Analysis of traffic accident severity using decision rules via decision trees,” Expert Syst. Appl., vol. 40, no. 15, pp. 6047–6054, 2013.
[64]J.AbelláN，G.LóPez和J. DeOñA，“通过决策树使用决策规则分析交通事故严重性”，Expert Syst。 Appl。，vol。 40，不。 15，pp.6047-6054,2013。

[65] C. Xie, J. Lu, and E. Parkany, “Work travel mode choice modeling with data mining: Decision trees and neural networks,” Transp. Res. Rec., J. Transp. Res. Board, pp. 50–61, Jan. 2003.
[65] C. Xie，J。Lu和E. Parkany，“使用数据挖掘的工作旅行模式选择建模：决策树和神经网络”，运输。 RES。 Rec。，J。Transp。 RES。董事会，第50-61页，2003年1月。

[66] E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, “Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach,” Transp. Res. C, Emerg. Technol., vol. 13, no. 3, pp. 211–234, 2005.
[66] E. I. Vlahogianni，M。G. Karlaftis和J. C. Golias，“用于短期交通流量预测的优化和元优化神经网络：遗传方法”，运输。 RES。 C，Emerg。 Technol。，vol。 13，不。 3，pp.211-234,2005。

[67] J. Van Lint, S. P. Hoogendoorn, and H. J. van Zuylen, “Accurate freeway travel time prediction with state-space neural networks under missing data,” Transp. Res. C, Emerg. Technol., vol. 13, nos. 5–6, pp. 347–369, Oct./Dec. 2005.
[67] J. Van Lint，S。P. Hoogendoorn和H. J. van Zuylen，“在缺失数据下使用状态空间神经网络进行精确的高速公路旅行时间预测”，运输。 RES。 C，Emerg。 Technol。，vol。 13，没有。 5-6，pp.347-369，Oct。/ Dec。2005年。

[68] X. Jin, R. L. Cheu, and D. Srinivasan, “Development and adaptation of constructive probabilistic neural network in freeway incident detection,” Transp. Res. C, Emerg. Technol., vol. 10, no. 2, pp. 121–147, 2002.
[68] X. Jin，R。L. Cheu和D. Srinivasan，“高速公路事件检测中建设性概率神经网络的发展和适应”，运输。 RES。 C，Emerg。 Technol。，vol。 10，不。 2，pp.121-147,2002。

[69] X. Zhu, J. Guo, and W. Huang, “Short-term forecasting of remaining parking spaces in parking guidance systems,” in Proc. 95th Annu. Meeting Transp. Res. Board, 2016.
[69] X. Zhu，J。Guo和W. Huang，“停车引导系统中剩余停车位的短期预测”，Proc。 95th Annu。会议运输。 RES。董事会，2016年。

[70] L. Vanajakshi and L. R. Rilett, “Support vector machine technique for the short term prediction of travel time,” in Proc. IEEE Intell. Veh. Symp., Jun. 2007, pp. 600–605.
[70] L. Vanajakshi和L. R. Rilett，“用于短期预测旅行时间的支持向量机技术”，Proc。 IEEE Intell。辆。 Symp。，Jun.2007，pp.600-605。

[71] Y. Bin, Y. Zhongzhen, and Y. Baozhen, “Bus arrival time prediction using support vector machines,” J. Intell. Transp. Syst., vol. 10, no. 4, pp. 151–158, 2006.
[71] Y. Bin，Y。Zhongzhen和Y. Baozhen，“使用支持向量机的公交车到达时间预测”，J。Intell。运输。 Syst。，vol。 10，不。 4，pp.151-158,2006。

[72] J. Xiao and Y. Liu, “Traffic incident detection using multiple-kernel support vector machine,” Transp. Res. Rec., J. Transp. Res. Board, pp. 44–52, Dec. 2012.

[72] J. Xiao和Y. Liu，“使用多核支持向量机的交通事件检测”，运输。 RES。 Rec。 ，J。Transp。 RES。董事会，第44-52页，2012年12月。

[73] Y. Meng and X. Liu, “Application of K-means algorithm based on ant clustering algorithm in macroscopic planning of highway transportation hub,” in Proc. 1st IEEE Int. Symp. Inf. Technol. Appl. Edu. (ISITAE), Nov. 2007, pp. 483–488.
[73] Y. Meng和X. Liu，“基于蚂蚁聚类算法的K-means算法在公路交通枢纽宏观规划中的应用”，Proc。第一届IEEE国际SYMP。天道酬勤。 TECHNOL。申请埃杜。 （ISITAE），2007年11月，第483-488页。

[74] R. P. D. Nath, H.-J. Lee, N. K. Chowdhury, and J.-W. Chang, “Modified k-means clustering for travel time prediction based on historical traffic data,” in Proc. Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst., 2010, pp. 511–521.
[74] R. P. D. Nath，H.-J。 Lee，N。K. Chowdhury和J.-W. Chang，“基于历史交通数据的旅行时间预测的修改k均值聚类”，Proc。诠释。 CONF。 Knowl.-Based Intell。天道酬勤。工程。 Syst。 ，2010，pp.511-521。

[75] B. Abdulhai, R. Pringle, and G. J. Karakoulas, “Reinforcement learning for true adaptive traffic signal control,” J. Transp. Eng., vol. 129, no. 3, pp. 278–285, 2003.
[75] B. Abdulhai，R。Pringle和G. J. Karakoulas，“加强学习实现真正的自适应交通信号控制”，J。运输。 Eng。 ，vol。 129，不。 3，pp.278-285,2003。

[76] I. Arel, C. Liu, T. Urbanik, and A. G. Kohls, “Reinforcement learningbased multi-agent system for network traffic signal control,” IET Intell. Transp. Syst., vol. 4, no. 2, pp. 128–135, 2010.
[76] I. Arel，C。Liu，T。Urbanik和A. G. Kohls，“基于强化学习的网络流量信号控制的多智能体系统”，IET Intell。运输。 Syst。 ，vol。 4，不。 2，pp.128-135,2010。

[77] A. L. C. Bazzan, “Opportunities for multiagent systems and multiagent reinforcement learning in traffic control,” Auto. Agents Multi-Agent Syst., vol. 18, no. 3, pp. 342–375, Jun. 2009.
[77] A. L. C. Bazzan，“多流量系统的机会和交通控制中的多代理强化学习”，Auto。Agent Multi-Agent Syst。 ，vol。 18，不。 3，pp.342-375，2009年6月。

[78] L. Li, Y. Lv, and F.-Y. Wang, “Traffic signal timing via deep reinforcement learning,” IEEE/CAA J. Automat. Sinica, vol. 3, no. 3, pp. 247–254, Apr. 2016.
[78] L. Li，Y。Lv和F.-Y. Wang，“Traf fi c信号时序通过深度强化学习”，IEEE / CAA J. Automat。 Sinica，vol。 3，不。 3，pp.247-254，2016年4月。

[79] X. Ma, H. Yu, Y. Wang, and Y. Wang, “Large-scale transportation network congestion evolution prediction using deep learning theory,” PLoS ONE, vol. 10, no. 3, p. e0119044, 2015.
[79] X. Ma，H。Yu，Y。Wang和Y. Wang，“使用深度学习理论的大规模交通网络拥挤演变预测”，PLoS ONE，vol。 10，不。 3，p。 e0119044,2015。

[80] H. Hu, B. Tang, X. Gong, W. Wei, and H. Wang, “Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks,” IEEE Trans. Ind. Informat., vol. 13, no. 4, pp. 2106–2116, Aug. 2017.
[80] H. Hu，B。Tang，X。Gong，W。Wei和H. Wang，“基于深度神经网络的大数据高速列车的智能故障诊断”，IEEE Trans。 Ind.Informat。 ，vol。 13，不。 4，pp.2106-2116，2017年8月。

[81] T. Chen, “Going deeper with convolutional neural network for intelligent transportation,” Ph.D. dissertation, Dept.

Elect. Comput. Eng., Worcester Polytech. Inst., Worcester, MA, USA, 2015.
[81] T. Chen，“深入研究卷积神经网络进行智能交通”，博士。论文，当选部门。 COMPUT。 Eng。 ，伍斯特 Polytech。 Inst。 ，Worcester，MA，USA，2015。

[82] Y. Duan, Y. Lv, W. Kang, and Y. Zhao, “A deep learning based approach for traffic data imputation,” in Proc. IEEE 17th Int. Conf. Intell. Transp. Syst. (ITSC), Oct. 2014, pp. 912–917.
[82] Y. Duan，Y。Lv，W。Kang和Y. Zhao，“基于深度学习的交通数据插补方法”，Proc。 IEEE 17th Int。 CONF。 INTELL。运输。 SYST。 （ITSC），2014年10月，第912-917页。

[83] N. Polson and V. Sokolov. (2016). “Deep learning for short-term traffic flow prediction.” [Online]. Available: https://arxiv.org/abs/1604.04527
[83] N. Polson和V. Sokolov。 （2016）。 “深入学习短期交通流量预测。”[在线]。可用：https：//arxiv.org/abs/1604.04527

[84] W. Huang, G. Song, H. Hong, and K. Xie, “Deep architecture for traffic flow prediction: Deep belief networks with multitask learning,” IEEE Trans. Intell. Transp. Syst., vol. 15, no. 5, pp. 2191–2201, Oct. 2014.
[84] W. Huang，G。Song，H。Hong和K. Xie，“交通流量预测的深层架构：具有多任务学习的深层信念网络”，IEEE Trans。 INTELL。运输。 Syst。 ，vol。 15，不。 5，pp.2191-2201，2014年10月。

[85] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, “Traffic flow prediction with big data: A deep learning approach,” IEEE Trans. Intell. Transp. Syst., vol. 16, no. 2, pp. 865–873, Apr. 2015.
[85] Y. Lv，Y。Duan，W。Kang，Z。Li和F.-Y. Wang，“Traf fi c fl ow预测大数据：深度学习方法，”IEEE Trans。 INTELL。运输。 Syst。 ，vol。 16，不。 2，pp.865-873，2015年4月。

[86] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, “Long short-term memory neural network for traffic speed prediction using remote microwave sensor data,” Transp. Res. C, Emerg. Technol., vol. 54, pp. 187–197, May 2015.
[86] X. Ma，Z。Tao，Y。Wang，H。Yu和Y. Wang，“使用远程微波传感器数据进行交通速度预测的长期短期记忆神经网络”，运输。 RES。 C，Emerg。 Technol。 ，vol。 54，pp.187-197，2015年5月。

[87] J. Zhai, Y. Cao, and Y. Chen, “Semantic information retrieval based on fuzzy ontology for intelligent transportation systems,” in Proc. IEEE Int. Conf. Syst., Man Cybern. (SMC), Oct. 2008, pp. 2321–2326.
[87] J. Zhai，Y。Cao和Y. Chen，“基于模糊本体的智能交通系统的语义信息检索”，Proc。 IEEE Int。 CONF。 Syst。 ，Man Cybern。 （SMC），2008年10月，第2321-2326页。

[88] S. Fernandez and T. Ito, “Driver behavior model based on ontology for intelligent transportation systems,” in Proc. IEEE 8th Int. Conf. Service-Oriented Comput. Appl. (SOCA), Oct. 2015, pp. 227–231.
[88] S. Fernandez和T. Ito，“基于智能交通系统本体的驾驶员行为模型”，Proc。 IEEE 8th Int。 CONF。面向服务的计算机。申请（SOCA），2015年10月，第227-231页。

[89] S. Fernandez and T. Ito, “Using SSN ontology for automatic traffic light settings on inteligent transportation systems,” in Proc. IEEE Int. Conf. Agents (ICA), Sep. 2016, pp. 106–107.
[89] S. Fernandez和T. Ito，“在智能交通系统中使用SSN本体进行自动交通灯设置”，Proc。 IEEE Int。 CONF。代理商（ICA），2016年9月，第106-107页。

[90] D. Gregor et al., “A methodology for structured ontology construction applied to intelligent transportation systems,”

Comput. Standards Interfaces, vol. 47, pp. 108–119, Aug. 2016.
[90] D. Gregor等人，“用于智能交通系统的结构化本体构建的方法论”，Comput。标准接口，第一卷。 47，pp.108-119，2016年8月。

[91] L. Zhao, R. Ichise, S. Mita, and Y. Sasaki, “Ontologies for advanced driver assistance systems,” J. Jpn. Soc. Artif. Intell., 2015, accessed: Aug. 12, 2016. [Online]. Available: http://www.ei.sanken.osakau.ac.jp/sigswo/papers/SIG-SWO-035/SIG-SWO-035-03.pdf
[91] L. Zhao，R。Ichise，S。Mita和Y. Sasaki，“Ontologies for advanced driver assistance systems，”J。Jpn。SOC。 ARTIF。 Intell。，2015，访问时间：2016年8月12日。[在线]。可用：http：//www.ei.sanken.osakau.ac.jp/sigswo/papers/SIG-SWO-035/SIG-SWO-035-03.pdf

[92] D. Chen, F. Asplund, K. Östberg, E. Brezhniev, and V. Kharchenko, “Towards an ontology-based approach to safety management in cooperative intelligent transportation systems,” in Proc. 10th Int. Conf. Depend. Complex Syst. Depcos-Relcomex, 2015, pp. 107–115.
[92] D. Chen，F。Asplund，K.Östberg，E。Brezhniev和V. Kharchenko，“迈向基于本体的合作智能交通系统安全管理方法”，Proc。第十届国际CONF。依靠。复杂的系统。 Depcos-Relcomex，2015，pp.107-115。

[93] W.-D. Yang and T. Wang, “The fusion model of intelligent transportation systems based on the urban traffic ontology,” Phys. Proc., vol. 25, no. 49, pp. 917–923, 2012.
[93] W.-D. Yang和T. Wang，“基于城市交通本体的智能交通系统融合模型”，Phys。 Proc。，vol。 25，不。 49，pp.917-923,2012。

[94] T. Toroyan, “Global status report on road safety,” Injury Prevention, vol. 15, no. 4, p. 286, 2009.
[94] T. Toroyan，“全球道路安全状况报告”，伤害预防，第一卷。 15，不。 4，p。 286,2009。

[95] T. F. Golob and W. W. Recker, “Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions,” J. Transp. Eng., vol. 129, no. 4, pp. 342–353, 2003.
[95] T. F. Golob和W. W. Recker，“城市高速公路事故，交通流量，天气和照明条件之间的关系”，J。运输。Eng。，vol。 129，不。 4，pp.342-353,2003。

[96] G. Xiong, F. Zhu, H. Fan, X. Dong, W. Kang, and T. Teng, “Novel ITS based on space-air-ground collected big-data,” in Proc. IEEE Int. Conf. Intell. Transp. Syst., Oct. 2014, pp. 1509–1514.
[96] G. Xiong，F。Zhu，H。Fan，X。Dong，W。Kang和T. Teng，“基于空间 - 地面收集大数据的新型ITS”，Proc。 IEEE Int。 CONF。 INTELL。运输。 Syst。，2014年10月，第1509-1514页。

[97] J. Lee and F. Mannering, “Impact of roadside features on the frequency and severity of run-off-roadway accidents: An empirical analysis,” Accident Anal. Prevention, vol. 34, no. 2, pp. 149–161, 2002.
[97] J. Lee和F. Mannering，“路边特征对径流事故频率和严重程度的影响：实证分析，”事故肛门。预防，第一卷34，不。 2，pp.149-161,2002。

[98] M. G. Karlaftis and I. Golias, “Effects of road geometry and traffic volumes on rural roadway accident rates,” Accident Anal. Prevention, vol. 34, no. 3, pp. 357–365, 2002.
[98] M. G. Karlaftis和I. Golias，“道路几何形状和交通量对农村道路事故率的影响”，事故肛门。预防，第一卷34，不。 3，pp.357-365,2002。

[99] L.-Y. Chang and W.-C. Chen, “Data mining of tree-based models to analyze freeway accident frequency,” J. Safety Res., vol. 36, no. 4, pp. 365–375, 2005.
[99] L.-Y. Chang和W.-C. Chen，“用于分析高速公路事故频率的树型模型的数据挖掘”，J。Safety Res。，vol。36，不。 4，pp.365-375,2005。

[100] M. Bédard, G. H. Guyatt, M. J. Stones, and J. P. Hirdes, “The independent contribution of driver, crash, and vehicle characteristics to driver fatalities,” Accident Anal. Prevention, vol. 34, no. 6, pp. 717–727, 2002.
[100]M.Bédard，G。H. Guyatt，M。J. Stones和J. P. Hirdes，“驾驶员，车祸和车辆特征对驾驶员死亡的独立贡献，”事故肛门。预防，第一卷34，不。 6，pp.717-727,2002。

[101] R. Li, C. Jiang, F. Zhu, and X. Chen, “Traffic flow data forecasting based on interval type-2 fuzzy sets theory,” IEEE/CAA J. Autom. Sinica, vol. 3, no. 2, pp. 141–148, Apr. 2016.
[101] R. Li，C。Jiang，F。Zhu和X. Chen，“基于区间类型2模糊集理论的交通流数据预测”，IEEE / CAA J. Autom。 Sinica，vol。 3，不。 2，pp.141-148，2016年4月。

[102] D. Chen, “Research on traffic flow prediction in the big data environment based on the improved RBF neural network,” IEEE Trans. Ind. Informat., vol. 13, no. 4, pp. 2000–2008, Aug. 2017.
[102] D. Chen，“基于改进的RBF神经网络的大数据环境中的交通流预测研究”，IEEE Trans。 Ind.Informat。，vol。 13，不。 4，pp.2000-2008，2017年8月。

[103] S. Jeon and B. Hong, “Monte Carlo simulation-based traffic speed forecasting using historical big data,” Future Generat. Comput. Syst., vol. 65, pp. 182–195, Dec. 2016.
[103] S. Jeon和B. Hong，“基于蒙特卡罗模拟的交通速度预测使用历史大数据，”Future Generat。 COMPUT。Syst。，vol。 65，pp.182-195，2016年12月。

[104] X.-L. Liu, P. Jia, S.-H. Wu, and B. Yu, “Short-term traffic flow forecasting based on multi-dimensional parameters,” J. Transp. Syst. Eng. Inf. Technol., vol. 11, no. 4, pp. 140–146, 2011.
[104] X.-L. Liu，P。Jia，S.-H。 Wu和B. Yu，“基于多维参数的短期交通流量预测”，J。运输。 SYST。工程。天道酬勤。 Technol。，vol。 11，不。 4，pp.140-146,2011。

[105] H.-H. Dong, X.-L. Sun, L.-M. Jia, H.-J. Li, and Y. Qin, “Traffic condition estimation with pre-selection space time model,” J. Central South Univ., vol. 19, no. 1, pp. 206–212, 2012.
[105] H.-H.董，X.-L.太阳，L.-M。贾，H.-J。 Li和Y. Qin，“具有预选空间时间模型的交通条件估计”，J。中南大学，第一卷。 19，没有。 1，pp.206-212,2012。

[106] M. Canaud, L. Mihaylova, J. Sau, and N.-E. El Faouzi, “Probabilty hypothesis density filtering for real-time traffic state estimation and prediction,” Netw. Heterogeneous Media, vol. 8, no. 3, pp. 825–842, 2013.
[106] M. Canaud，L。Mihaylova，J。Sau和N.-E. El Faouzi，“用于实时交通状态估计和预测的概率假设密度滤波”，Netw。异构媒体，第一卷。 8，不。 3，pp.825-842,2013。

[107] T. L. Pan, A. Sumalee, R. X. Zhong, and N. Indra-Payoong, “Short-term traffic state prediction based on temporal–spatial correlation,” IEEE Trans. Intell. Transp. Syst., vol. 14, no. 3, pp. 1242–1254, Sep. 2013.
[107] T. L. Pan，A。Sumalee，R。X. Zhong和N. Indra-Payoong，“基于时空相关的短期交通状态预测”，IEEE Trans。 INTELL。运输。 Syst。，vol。 14，没有。 3，pp.1242-1254，2013年9月。

[108] C. Antoniou, H. N. Koutsopoulos, and G. Yannis, “Dynamic datadriven local traffic state estimation and prediction,” Transp. Res. C, Emerg. Technol., vol. 34, pp. 89–107, Sep. 2013.
[108] C. Antoniou，H。N. Koutsopoulos和G. Yannis，“动态数据驱动局部交通状态估计和预测”，运输。 RES。 C，Emerg。 Technol。，vol。 34，pp.89-107，2013年9月。

[109] B. Ghosh, B. Basu, and M. O’Mahony, “Multivariate short-term traffic flow forecasting using time-series analysis,” IEEE Trans. Intell. Transp. Syst., vol. 10, no. 2, pp. 246–254, Jun. 2009.
[109] B. Ghosh，B。Basu和M. O'Mahony，“使用时间序列分析的多变量短期交通流量预测”，IEEE Trans。 INTELL。运输。 Syst。，vol。 10，不。 2，pp.246-254，2009年6月。

[110] J. Xu, D. Deng, U. Demiryurek, C. Shahabi, and M. van der Schaar, “Mining the situation: Spatiotemporal traffic prediction with big data,” IEEE J. Sel. Topics Signal Process., vol. 9, no. 4, pp. 702–715, Jun. 2015.
[110] J. Xu，D。Deng，U。Demiryurek，C。Shahabi和M. van der Schaar，“挖掘情况：利用大数据进行时空交通预测”，IEEE J. Sel。主题信号处理。，第一卷。 9，不。 4，pp.702-715，2015年6月。

[111] H.-P. Lu, Z.-Y. Sun, and W.-C. Qu, “Big data and its applications in urban intelligent transportation system,” J. Transp. Syst. Eng. Inf. Technol., vol. 15, no. 5, pp. 45–52, 2015.
[111] H.-P. Lu，Z.-Y。 Sun和W.-C. Qu，“大数据及其在城市智能交通系统中的应用”，J。运输。 SYST。工程。天道酬勤。 Technol。，vol。 15，不。 5，pp.45-52,2015。

[112] C.-C. Lu, X. Zhou, and K. Zhang, “Dynamic origin–destination demand flow estimation under congested traffic conditions,” Transp. Res. C, Emerg. Technol., vol. 34, pp. 16–37, Sep. 2013.
[112] C.-C. Lu，X。Zhou和K. Zhang，“在拥挤的交通条件下动态起源 - 目的地需求流量估算”，运输。 RES。 C，Emerg。 Technol。，vol。 34，pp.16-37，2013年9月。

[113] L. Alexander, S. Jiang, M. Murga, and M. C. González, “Origin– destination trips by purpose and time of day inferred from mobile phone data,” Transp. Res. C, Emerg. Technol., vol. 58, pp. 240–250, Sep. 2015.
[113] L. Alexander，S。Jiang，M。Murga和M.C.González，“根据目的和时间从手机数据推断的目的地旅行”，运输。 RES。 C，Emerg。 Technol。，vol。 58，pp.240-250，2015年9月。

[114] J. B. Gordon, “Intermodal passenger flows London’s public transport network: Automated inference full passenger journeys using faretransaction vehicle-location data,” Ph.D. dissertation, Dept. Urban Studies Planning, Dept. Civil Environ. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 2012.
[114] J. B. Gordon，“多式联运乘客飞行伦敦的公共交通网络：利用交通车辆位置数据自动推断全程乘客旅程”，博士。学位论文，城市研究规划系，民间环境系。 Eng。，Massachusetts Inst。 Technol。，Cambridge，MA，USA，2012。

[115] S. Tao, “Investigating the travel behaviour dynamics of bus rapid transit passengers,” Ph.D. dissertation, School Geograp., Planning Environ. Manage., Univ. Queensland, Brisbane, Qld, Australia, 2015.
[115] S. Tao，“调查公交捷运乘客的旅行行为动态”，博士。学位论文，学术地理学，规划环境。管理。，大学昆士兰，布里斯班，昆士兰，澳大利亚，2015年。

[116] I. Gokasar and K. Simsek, “Using ‘Big data’ for analysis and improvement of public transportation systems in Istanbul,” Tech. Rep., 2014.

[116] I. Gokasar和K. Simsek，“利用'大数据'分析和改善伊斯坦布尔的公共交通系统，”Tech。 2014年代表。

[117] J. Chan et al., “Rail transit OD matrix estimation and journey time reliability metrics using automated fare data,” PhD thesis, Dept. Civil Environ. Eng., Massachusetts Inst. Technol., Cambridge, MA, USA, 2007.
[117] J. Chan等人，“使用自动票价数据的轨道交通OD矩阵估计和旅程时间可靠性指标”，博士论文，民用环境部。 Eng。 ，Massachusetts Inst。 Technol。 ，Cambridge，MA，USA，2007。

[118] J. L. Toole, S. Colak, B. Sturt, L. P. Alexander, A. Evsukoff, and M. C. González, “The path most traveled: Travel demand estimation using big data resources,” Transp. Res. C, Emerg. Technol., vol. 58, pp. 162–177, Sep. 2015.
[118] J. L. Toole，S。Colak，B。Sturt，L。P. Alexander，A。Evsukoff和M.C.González，“最常旅行的路径：使用大数据资源进行旅行需求估算”，运输。 RES。 C，Emerg。 Technol。 ，vol。 58，pp.162-177，2015年9月。

[119] B. Ferris, K. Watkins, and A. Borning, “OneBusAway: Results from providing real-time arrival information for public transit,” in Proc. SIGCHI Conf. Human Factors Comput. Syst., 2010, pp. 1807–1816. [120] [Online]. Available: http://inrix.com/mobile-apps/ [121] [Online]. Available: https://www.waze.com/ [122] [Online]. Available: http://moovitapp.com/ [123] [Online]. Available: http://gaode.com/ [124] “Opening up to open data, ” in Proc. Int. Assoc. Public Transp., 2014.
[119] B. Ferris，K。Watkins和A. Borning，“OneBusAway：为公共交通提供实时到达信息的结果”，Proc。 SIGCHI Conf。人为因素计算Syst。 ，2010，pp.1807-1816。 [120] [在线]。可用：http：//inrix.com/mobile-apps/ [121] [在线]。可用：https：//www.waze.com/ [122] [在线]。可用：http：//moovitapp.com/ [123] [在线]。可用：http：//gaode.com/ [124]“打开数据开放”，在Proc。诠释。协会。公共运输，2014年。

[125] B. Schultz, Ed., “Operational analytics keeps bay area trains on track,” All Analytics, May 2012. [Online]. Available: http:// www.allanalytics.com/author.asp?section_id=1411&doc_id=244062
[125] B. Schultz，Ed。 ，“运营分析使海湾地区列车保持正常”，All Analytics，2012年5月。[在线]。可用：http：//www.allanalytics.com/author.asp？section_id = 1411＆doc_id = 244062

[126] M. Faizrahnemoon, A. Schlote, L. Maggi, E. Crisostomi, and R. Shorten, “A big-data model for multi-modal public transportation with application to macroscopic control and optimisation,” Int. J. Control, vol. 88, no. 11, pp. 2354–2368, 2015.
[126] M. Faizrahnemoon，A。Schlote，L。Maggi，E。Crisostomi和R. Shorten，“用于宏观控制和优化的多模式公共交通的大数据模型”，Int。 J. Control，vol。 88，不。 11，pp.2354-2368,2015。

[127] N. Van Oort, “Big data opportunities in public transport: Enhancing public transport by ITCS,” in Proc. IT-TRANS, Karlsruhe, Germany, Feb. 2014.
[127] N. Van Oort，“公共交通中的大数据机遇：通过ITCS加强公共交通”，Proc。 IT-TRANS，卡尔斯鲁厄，德国，2014年2月。

[128] Z. Jiang, C.-H. Hsu, D. Zhang, and X. Zou, “Evaluating rail transit timetable using big passengers’ data,” J. Comput. Syst. Sci., vol. 82, no. 1, pp. 144–155, 2016.
[128] Z. Jiang，C.-H。 Hsu，D。Zhang和X. Zou，“使用大乘客数据评估轨道交通时刻表”，J。Comput。 SYST。 Sci。 ，vol。 82，不。 1，pp.144-155,2016。

[129] J. Yin, D. Chen, and Y. Li, “Smart train operation algorithms based on expert knowledge and ensemble CART for

the electric locomotive,” Knowl.-Based Syst., vol. 92, pp. 78–91, Jan. 2016.
[129] J. Yin，D。Chen和Y. Li，“基于专家知识的智能列车运行算法和用于电力机车的集合CART”，基于Knowl.-Based Syst。，vol。 92，pp.78-91，2016年1月。

[130] D. Chen, T. Tang, C. Gao, and R. Mu, “Research on the error estimation models and online learning algorithms for train station parking in urban rail transit,” China Railway Sci., vol. 31, no. 6, pp. 122–127, 2010.
[130] D. Chen，T。Tang，C。Gao和R. Mu，“关于城市轨道交通中火车站停车的误差估计模型和在线学习算法的研究”，中国铁道科学，第一卷。 31，不。 6，pp.122-127,2010。

[131] J. Zhou, “Applications of machine learning methods in problem of precise train stopping,” Comput. Eng. Appl., vol. 46, no. 25, pp. 226–230, 2010.
[131] J. Zhou，“机器学习方法在精确列车停止问题中的应用”，Comput。工程。 Appl。，vol。 46，不。 25，pp.226-230,2010。

[132] Z. Hou, Y. Wang, C. Yin, and T. Tong, “Terminal iterative learning control based station stop control of a train,” Int. J. Control, vol. 84, no. 7, pp. 1263–1274, 2011.
[132] Z. Hou，Y。Wang，C。Yin和T. Tong，“基于终端迭代学习控制的火车站停止控制”，Int。 J. Control，vol。 84，不。 7，pp.1263-1274,2011。

[133] D. Chen, R. Chen, T. Tang, and Y. Li, “Online learning algorithms for train automatic stop control using precise location data of balises,” IEEE Trans. Intell. Transp. Syst., vol. 14, no. 3, pp. 1526–1535, Sep. 2013.
[133] D. Chen，R。Chen，T。Tang和Y. Li，“使用精确的应答器位置数据进行列车自动停止控制的在线学习算法”，IEEE Trans。 INTELL。运输。 Syst。，vol。 14，没有。 3，pp.1526-1535，2013年9月。

[134] A. Thaduri, D. Galar, and U. Kumar, “Railway assets: A potential domain for big data analytics,” Proc. Comput. Sci., vol. 53, no. 1, pp. 457–467, 2015.
[134] A. Thaduri，D。Galar和U. Kumar，“铁路资产：大数据分析的潜在领域”，Proc。 COMPUT。 Sci。，vol。 53，不。 1，pp.457-467,2015。

[135] A. Núñez, J. Hendriks, Z. Li, B. De Schutter, and R. Dollevoet, “Facilitating maintenance decisions on the dutch railways using big data: The aba case study,” in Proc. IEEE Int. Conf. Big Data, Oct. 2014, pp. 48–53.
[135]A.Núñez，J。Hendriks，Z。Li，B。De Schutter和R. Dollevoet，“利用大数据促进荷兰铁路的维护决策：aba案例研究”，Proc。 IEEE Int。 CONF。大数据，2014年10月，第48-53页。

[136] J. Tutcher, “Ontology-driven data integration for railway asset monitoring applications,” in Proc. IEEE Int. Conf. Big Data, Oct. 2014, pp. 85–95.
[136] J. Tutcher，“用于铁路资产监测应用的本体驱动的数据集成”，Proc。 IEEE Int。 CONF。大数据，2014年10月，第85-95页。

[137] A. M. Zarembski, “Some examples of big data in railroad engineering,” in Proc. IEEE Int. Conf. Big Data, Oct. 2014, pp. 96–102.
[137] A. M. Zarembski，“铁路工程大数据的一些例子”，Proc。 IEEE Int。 CONF。大数据，2014年10月，第96-102页。

[138] Asset Management Services, Network Rail, London, U.K., 2013.

[138]资产管理服务，网络铁路，伦敦，英国，2013年。

[139] H. Li et al., “Improving rail network velocity: A machine learning approach to predictive maintenance,” Transp. Res. C, Emerg. Technol., vol. 45, pp. 17–26, Aug. 2014.
[139] H. Li等人，“改善铁路网速度：机器学习方法进行预测性维护”，运输。 RES。 C，Emerg。 Technol。 ，vol。 45，第17-26页，2014年8月。

[140] F. Wang, T. Xu, T. Tang, M. Zhou, and H. Wang, “Bilevel feature extraction-based text mining for fault diagnosis of railway systems,” IEEE Trans. Intell. Transp. Syst., vol. 18, no. 1, pp. 49–58, Jan. 2017.
[140] F. Wang，T。Xu，T。Tang，M。Zhou和H. Wang，“基于双层特征提取的铁路系统故障诊断文本挖掘”，IEEE Trans。 INTELL。运输。 Syst。 ，vol。 18，不。 1，pp.49-58，2017年1月。

[141] X. Meng et al., “Mllib: Machine learning in apache spark,” J. Mach. Learn. Res., vol. 17, no. 34, pp. 1–7, 2016.
[141] X. Meng等人，“Mllib：apache spark中的机器学习”，J。Mach。学习。 Res。 ，vol。 17，不。 34，第1-7页，2016年。

[142] R. Mian, H. Ghanbari, S. Zareian, M. Shtern, and M. Litoiu, “A data platform for the highway traffic data,” in Proc. MESOCA, 2014, pp. 47–52.
[142] R. Mian，H.Ghanbari，S。Zareian，M。Shtern和M. Litoiu，“高速公路交通数据的数据平台”，Proc。 MESOCA，2014年，第47-52页。

[143] S. Zareian, R. Veleda, M. Litoiu, M. Shtern, H. Ghanbari, and M. Garg, “K-feed—A data-oriented approach to application performance management in cloud,” in Proc. IEEE 8th Int. Conf. Cloud Comput., Jun. 2015, pp. 1045–1048.
[143] S. Zareian，R。Veleda，M。Litoiu，M。Shtern，H. Ghanbari和M. Garg，“K-feed-A数据导向的云应用程序性能管理方法”，Proc。 IEEE 8th Int。 CONF。 Cloud Comput。 ，2015年6月，第1045-1048页。

[144] M. Shtern, R. Mian, M. Litoiu, S. Zareian, H. Abdelgawad, and A. Tizghadam, “Towards a multi-cluster analytical engine for transportation data,” in Proc. Int. Conf. Cloud Auton. Comput. (ICCAC), 2014, pp. 249–257.
[144] M. Shtern，R。Mian，M。Litoiu，S。Zareian，H。Abdelgawad和A. Tizghadam，“迈向交通数据的多集群分析引擎”，Proc。诠释。 CONF。 Cloud Auton。 COMPUT。 （ICCAC），2014年，第249-257页。

[145] H. Khazaei, S. Zareian, R. Veleda, and M. Litoiu, “Sipresk: A big data analytic platform for smart transportation,” in Proc. EAI Int. Conf. Big Data Anal. Smart Cities, 2015, pp. 419–430.
[145] H. Khazaei，S。Zareian，R。Veleda和M. Litoiu，“Sipresk：智能交通的大数据分析平台”，Proc。 EAI Int。 CONF。大数据肛门。智慧城市，2015年，第419-430页。

[146] J. Chaolong, H. Wang, and L. Wei, “Study of smart transportation data center virtualization based on vmware vsphere and parallel continuous query algorithm over massive data streams,” Proc. Eng., vol. 137, no. 6, pp. 719–728, 2016.
[146] J. Chaolong，H。Wang和L. Wei，“研究基于vmware vsphere的智能交通数据中心虚拟化和基于海量数据流的并行连续查询算法”，Proc。 Eng。 ，vol。 137，不。 6，pp.719-728,2016。

[147] I. R. Kamel, H. Abdelgawad, and B. Abdulhai, “Transportation big data simulation platform for the greater toronto area (GTA),” in Smart City 360°. New York, NY, USA: Springer, 2016, pp. 443–454.
[147] I. R. Kamel，H。Abdelgawad和B. Abdulhai，“更大多伦多地区的交通大数据模拟平台（GTA）”，在Smart

City 360°。纽约，纽约，美国：Springer，2016年，第443-454页。

[148] G. Guerreiro, P. Figueiras, R. Silva, R. Costa, and R. Jardim-Goncalves, “An architecture for big data processing on intelligent transportation systems: An application scenario on highway traffic flows,” in Proc. IEEE 8th Int. Conf. Intell. Syst. (IS), Sep. 2016, pp. 65–72.
[148] G. Guerreiro，P。Figueiras，R。Silva，R。Costa和R. Jardim-Goncalves，“智能交通系统大数据处理架构：高速公路交通流量的应用场景”，Proc。 IEEE 8th Int。 CONF。 INTELL。 SYST。 （IS），2016年9月，第65-72页。

[149] E. Bouillet et al., “Data stream processing infrastructure for intelligent transport systems,” in Proc. IEEE 66th Veh. Technol. Conf. (VTC-Fall), Oct. 2007, pp. 1421–1425.
[149] E. Bouillet等人，“智能交通系统的数据流处理基础设施”，Proc。 IEEE 66th Veh。 TECHNOL。 CONF。 （VTC-Fall），2007年10月，第1421-1425页。

[150] S. Amini, I. Gerostathopoulos, and C. Prehofer, “Big data analytics architecture for real-time traffic control,” in Proc. 5th IEEE Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS), Jun. 2017, pp. 710–715.
[150] S. Amini，I。Gerostathopoulos和C. Prehofer，“用于实时交通控制的大数据分析架构”，Proc。第五届IEEE国际CONF。型号Technol。 INTELL。运输。 SYST。 （MT-ITS），2017年6月，第710-715页。

[151] G. Zeng, “Application of big data in intelligent traffic system,” IOSR J. Comput. Eng., vol. 17, no. 1, pp. 1–4, 2015.
[151] G. Zeng，“大数据在智能交通系统中的应用”，IOSR J. Comput。 Eng。 ，vol。 17，不。 1，pp.1-4,2015。

[152] M. Tahmassebpour and A. M. Otaghvari, “Increase efficiency big data in intelligent transportation system with using IoT integration cloud,” J. Fundam. Appl. Sci., vol. 8, no. 3S, pp. 2443–2461, 2016.
[152] M. Tahmassebpour和A. M. Otaghvari，“使用物联网集成云提高智能交通系统的效率大数据”，J。Fundam。申请Sci。 ，vol。 8，不。 3S，pp.2443-2461,2016。

[153] M. Chowdhury, A. Apon, and K. Dey, Data Analytics for Intelligent Transportation Systems. Amsterdam, The Netherlands: Elsevier, 2017.
[153] M. Chowdhury，A。Apon和K. Dey，智能交通系统数据分析。荷兰阿姆斯特丹：Elsevier，2017。

[154] Z. Ji, I. Ganchev, M. O’Droma, L. Zhao, and X. Zhang, “A cloudbased car parking middleware for IoT-based smart cities: Design and implementation,” Sensors, vol. 14, no. 12, pp. 22372–22393, 2014.
[154] Z. Ji，I。Ganchev，M。O'Droma，L。Zhao和X. Zhang，“基于物联网的智能城市的基于云的停车中间件：设计和实施，”传感器，第一卷。 14，没有。 12，pp.22372-22393,2014。

[155] M. Smith, C. Szongott, B. Henne, and G. von Voigt, “Big data privacy issues in public social media,” in Proc. IEEE Int. Conf. Digit. Ecosyst. Technol., Jun. 2012, pp. 1–6.
[155] M. Smith，C。Szongott，B。Henne和G. von Voigt，“公共社交媒体中的大数据隐私问题”，Proc。 IEEE Int。 CONF。数字。 Ecosyst。 Technol。 ，2012年6月，第1-6页。

[156] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, “Enabling public auditability and data dynamics for storage security in cloud computing,” IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 5, pp. 847–859, May 2011.
[156] Q. Wang，C。Wang，K。Ren，W。Lou和J. Li，“为云计算中的存储安全性提供公共可审计性和数据动

态”，IEEE Trans。并行分销。 Syst。 ，vol。 22，不。 5，pp.847-859，2011年5月。

[157] O. Tene and J. Polonetsky, “Big data for all: Privacy and user control in the age of analytics,” Northwestern J. Technol. Intell. Property, vol. 11, no. 5, 2012, Art. no. 1.
[157] O. Tene和J. Polonetsky，“所有人的大数据：分析时代的隐私和用户控制”，Northwestern J. Technol。 INTELL。财产，第一卷11，不。 5,2012，Art。没有。 1。

[158] M. Hilbert and P. López, “The world’s technological capacity to store, communicate, and compute information,” Sci., vol. 332, no. 6025, pp. 60–65, 2011.
[158] M. Hilbert和P.López，“世界储存，交流和计算信息的技术能力”，Sci。 ，vol。 332，没有。 6025，pp.60-65,2011。

[159] M. D. Assunção, R. N. Calheiros, S. Bianchi, M. A. S. Netto, and R. Buyya, “Big Data computing and clouds: Trends and future directions,” J. Parallel Distrib. Comput., vols. 79–80, pp. 3–15, May 2013.
[159] M.D.Assunção，R。N. Calheiros，S。Bianchi，M。A. S. Netto和R. Buyya，“大数据计算和云：趋势和未来方向”，J。Parallel Distrib。 Comput。 ，vols。 79-80，第3-15页，2013年5月。

[160] J. Liu, J. Li, W. Li, and J. Wu, “Rethinking big data: A review on the data quality and usage issues,” ISPRS J. Photogram. Remote Sens., vol. 115, pp. 134–142, May 2016.
[160] J. Liu，J。Li，W。Li和J. Wu，“重新思考大数据：关于数据质量和使用问题的评论”，ISPRS J. Photogram。遥感。 ，第一卷。 115，pp.134-142，2016年5月。

[161] J. Li and X. Liu, “An important aspect of big data: Data usability,” J. Comput. Res. Develop., vol. 50, no. 6, pp. 1147–1162, 2013.
[161] J. Li和X. Liu，“大数据的一个重要方面：数据可用性，”J. Comput。 RES。发展。 ，第一卷。 50，不。 6，pp.1147-1162,2013。

Li Zhu received the Ph.D. degree in traffic control and information engineering from Beijing Jiaotong University, Beijing, China, in 2012. He is currently a Faculty Member at Beijing Jiaotong University and a Visiting Scholar at Carleton University, Ottawa, ON, Canada, and The University of British Columbia, Vancouver, BC, Canada. His research interests include intelligent transportation systems, train-ground communication technology in communication base train ground communication systems, and cross layer design in train-ground communication systems.
李朱获得博士学位。 2012年，北京交通大学交通控制与信息工程专业。他目前是北京交通大学的教员和加拿大安大略省渥太华卡尔顿大学的访问学者，以及加拿大不列颠哥伦比亚省温哥华的大学。他的研究兴趣包括智能交通系统，通信基地列车地面通信系统中的列车 - 地面通信技术，以及列车 - 地面通信系统中的跨层设计。

**Fei Richard Yu** (S'00–M'04–SM'08–F'18) received the Ph.D. degree in electrical engineering from The University of British Columbia in 2003. From 2002 to 2006, he was with Ericsson, Lund, Sweden, and a start-up in California, USA. He joined Carleton University in 2007, where he is currently a Professor. His research interests include wireless cyber-physical systems, connected/autonomous vehicles, security, distributed ledger technology, and deep learning. He received the IEEE Outstanding Service Award in 2016, the IEEE Outstanding Leadership Award in 2013, the Carleton Research Achievement Award in 2012, the Ontario Early Researcher Award (formerly Premiers Research Excellence Award) in 2011, the Excellent Contribution Award at IEEE/IFIP TrustCom 2010, the Leadership Opportunity Fund Award from Canada Foundation of Innovation in 2009, and the Best Paper Awards at IEEE VTC 2017 Spring, ICC 2014, Globecom 2012, IEEE/IFIP TrustCom 2009, and International Conference on Networking 2005.

Dr. Yu is a registered Professional Engineer in ON, Canada, and a fellow of the Institution of Engineering and Technology. He has served as the Technical Program Committee Co-Chair of numerous conferences. He serves on the Editorial Boards of several journals, including the Co-Editor-in-Chief for *Ad Hoc and Sensor Wireless Networks*, a Lead Series Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING, and IEEE COMMUNICATIONS SURVEYS AND TUTORIALS. He is a Distinguished Lecturer, the Vice President (Membership), and an Elected Member of the Board of Governors of the IEEE Vehicular Technology Society.

**Yige Wang** received the bachelor's degree in electrical engineering from Shandong University of Technology, Shandong, China, in 2015. He is currently pursuing the degree with Beijing Jiaotong University. His research interests include train-ground communication technology in communication base train ground communication systems and intelligent transportation systems.

**Bin Ning** (F'14) received the master's and Ph.D. degrees in engineering from Northern Jiaotong University in 1987 and 2005, respectively. He was a Visiting Scholar in electronics and electrical power engineering with Brunel University, London, U.K. From 2002 to 2003, he was with UC Berkeley as a Senior Visiting Scholar. He is currently a Professor with Beijing Jiaotong university. His research mainly focus on high speed train control system and railway transportation train control system, including main locomotive signal, communication based train control system, intelligent transportation, fault-tolerant design of signal system, fault diagnosis, system reliability, and security design. He is responsible for several key scientific and technical projects, and made great research achievement.

Dr. Ning is a member of the China Overseas Returned Scholars Association and a member of the Editorial Board of the Journal of Railways in China. He is a fellow of the Association of International Railway Signaling Engineers, The Institute of Engineering and Technology, and the China Railway Society. He is the Deputy Director of the China Traffic System Engineering Society and the Beijing Railway Society. He is the Chair of Technical Committee on Railroad Systems and Applications of the IEEE Intelligent Transportation Systems Society. He was an Associate Editor of IEEE TRANSACTION ON INTELLIGENT TRANSPORTATION SYSTEMS (2010–2012) and *Acta Automatica Sinica* (2011–2012).

Tao Tang received the Ph.D. degree in engineering from Chinese Academy of Science in 1991. He is currently a Professor with Beijing Jiaotong University and an Associate Director of the Rail Traffic Control and Safety State Key Laboratory. His research interests include communication based train control, high speed train control system, and intelligent transportation system.

陶唐获得博士学位。 1991年获得中国科学院工程学士学位。他目前是北京交通大学教授，铁路交通控制与安全国家重点实验室副主任。他的研究兴趣包括基于通信的列车控制，高速列车控制系统和智能交通系统。

Dr. Tang is a member of experts Group of High Technology Research and Development Program of China (863 Program) and the Leader in the Field of Modern Transportation Technology Experts Group. He is also a Specialist of National Development and Reform Commission and Beijing Urban Traffic Construction Committee.

唐博士是中国高技术研究与发展计划（863计划）专家组和现代交通技术专家组领导人。他还是国家发展和改革委员会和北京城市交通建设委员会的专家。